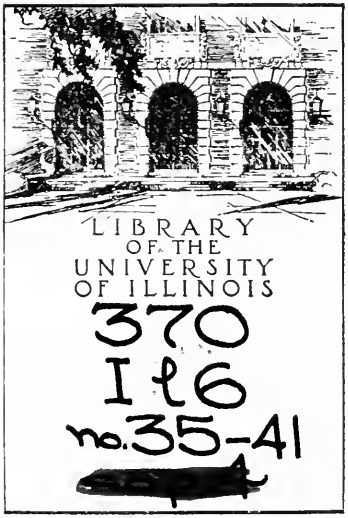


BULLETIN NO. 40

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

A GLOSSARY OF THREE HUNDRED TERMS USED IN EDUCATIONAL MEASUREMENT AND RESEARCH

By

CHARLES W. ODELL
Assistant Director, Bureau of Educational Research

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1928

ERRATA

Bureau of Educational Research, Bulletin No. 40

Page 9, Line 12. For "his" read "this."

Page 20, Line 33. Omit comma after "ordinary."

Page 29, Line 4. For "Q." read "Q_1."

Page 29, Line 7. The formula should read $Q_1 = l + \frac{\frac{N}{4} - S}{f} i.$

Page 33, Line 23. For "M" read "M_g."

Page 38, Line 35. The formula should read $Md. = l + \frac{\frac{N}{2} - S}{f} i.$

Page 44, Line 1. Omit "and at."

Page 45, Line 29. For "positive" read "partial."

Page 60, Line 25. The formula should read $\sigma = \sqrt{\frac{\Sigma x^2}{N}}.$

Page 63, Line 10. For "Q^3." read "Q_3"

Page 63, Line 13. The formula should read $Q_3 = l + \frac{\frac{3N}{4} - S}{f} i.$

The errors listed above are those which might easily mislead readers. Minor errors such as the misspelling of words, the insertion of periods following certain abbreviations where they are not commonly employed, and the omission of periods where required for purposes of punctuation, are not listed and corrected because they do not appear to offer opportunities for misunderstanding.

PREFACE

Circular Number 13, of the Bureau of Educational Research, which bore the title, "Definitions of the Terminology of Educational Measurements," is now out of print. The present bulletin is a revision and enlargement of this original publication. Practically all of the original definitions have been rewritten and references have been inserted so that one who desires further information can easily locate it.

Educational research, like many other fields of human endeavor, has a technical vocabulary. Many of the words and phrases included in it are also used in non-technical fields or even in ordinary communication. Whenever a word or phrase is used in a technical sense it has a very precise and definite meaning, which is usually not true in the case of its more popular usage. Consequently, it is highly important that one who is engaged in educational research, or one who reads reports of research, know the technical meanings of the words or phrases commonly used in this field.

WALTER S. MONROE, *Director.*

November 22, 1927.

A Glossary of Three Hundred Terms Used In Educational Measurement and Research

The terms defined or explained in this glossary were secured by the examination of some fifteen of the best and most widely used books in the general field covered, also of a number of articles in educational periodicals and of various other sources. As a result a list of about three hundred terms, not including abbreviations, which seemed to merit inclusion in such a publication as this was compiled. These were taken from both educational research in general and that dealing with tests and measurements of pupil ability and achievement. No texts in educational statistics were consulted, but because of the frequent use of statistical expressions in the field of measurements, a large number of such terms are contained in this glossary. Terms peculiar to research in lines other than tests and measurements, such as school buildings, finances, methods of teaching, the curriculum, and so forth, were not included, nor were those that may be classed as belonging to psychology rather than to education.

In such a list of terms there are, of course, many that are synonymous. In such instances the term most commonly used or preferred by the writer has been defined and the others given as synonymous with it. Such abbreviations as are commonly used in connection with any of the expressions in the list are given and referred to the proper terms. In many cases from one to three references have been given which may be consulted by readers who wish a more complete discussion than is contained in this publication. In some cases these references contain fuller definitions and explanations, in others examples and illustrations, and in others more general discussions of the use of the term defined. No attempt has been made to refer to original sources, nor have any periodical articles been mentioned. It seemed that if the references were limited to a dozen or so fairly well-known books and a very few other easily available publications, they would be more helpful and usable to the ordinary reader. Therefore this principle has been applied in the selection of references. To economize space the references in the text are limited to the name of the author and the pages, or in the case of two or more books by the same author, enough of the title to make clear which one is meant. The following is a complete list of the references mentioned:

FREEMAN, F. N. *Mental Tests.* Boston: Houghton Mifflin Company, 1926. 503 p.

KELLEY, T. L. *Interpretation of Educational Measurements.* Yonkers: World Book Company, 1927. 363 p.

MCCALL, W. A. *How to Experiment in Education.* New York: The Macmillan Company, 1923. 281 p.

MCCALL, W. A. *How to Measure in Education.* New York: The Macmillan Company, 1922. 416 p.

MONROE, W. S. "The Constant and Variable Errors of Educational Measurements," *University of Illinois Bulletin,* Vol. 21, No. 10. Bureau of Educational Research Bulletin No. 15. Urbana: University of Illinois, 1923. 30 p.

MONROE, W. S. *An Introduction to the Theory of Educational Measurements.* Boston: Houghton Mifflin Company, 1923. 364 p.

MONROE, W. S., DEVOSS, J. C., and KELLY, F. J. *Educational Tests and Measurements,* Revised and Enlarged Edition. Boston: Houghton Mifflin Company, 1924. 521 p.

MONROE, W. S. and ENGELHART, M. D. "The Techniques of Educational Research," *University of Illinois Bulletin,* Vol. 25, No. 19. Bureau of Educational Research Bulletin No. 38, Urbana: University of Illinois, 1928. 84 p.

ODELL, C. W. *Educational Statistics.* New York: Century Company, 1925. 334 p.

ODELL, C. W. "The Interpretation of the Probable Error and the Coefficient of Correlation," *University of Illinois Bulletin,* Vol. 23, No. 52. Bureau of Educational Research Bulletin No. 32. Urbana: University of Illinois, 1926. 49 p.

ODELL, C. W. "Objective Measurement of Information," *University of Illinois Bulletin,* Vol. 23, No. 36. Bureau of Educational Research Circular No. 44. Urbana: University of Illinois, 1926. 27 p.

OTIS, A. S. *Statistical Method in Educational Measurement.* Yonkers: World Book Company, 1925. 337 p.

RUCH, G. M. and STODDARD, G. D. *Tests and Measurements in High School Instruction.* Yonkers: World Book Company, 1927. 381 p.

RUGG, H. O. *Statistical Methods Applied to Education.* Boston: Houghton Mifflin Company, 1917. 410 p.

RUSSELL, CHARLES. *Classroom Tests.* Boston: Ginn and Company, 1926. 346 p.

SYMONDS, P. M. *Measurement in Secondary Education.* New York: The Macmillan Company, 1927. 588 p.

A. A. Abbreviation for *achievement age,* also *accomplishment age* and *attainment age.*

Accidental error. Synonymous with *variable error.*

Accomplishment age (A. A.) Sometimes used as synonymous with *achievement age.*

Accomplishment quotient (A. Q.) Sometimes used as synonymous with *achievement quotient.*

Accomplishment ratio (A. R.) A rarely employed term, synonymous with *achievement ratio.*

Accuracy. Accuracy refers in a general way to freedom from error. The term has two more or less special or technical uses in the field of educational measurement. In one of these it refers to a characteristic or dimension of pupil achievement and in this sense is very nearly synonymous with *quality.* It is, however, slightly more restricted in its meaning than quality and may be defined as the correctness or freedom from error of pupils' responses. In its second sense it is employed in connection with the freedom from error of test scores and other measures. In this connection it is sometimes used as synonymous with *reliability,* but really has a broader meaning since reliability is concerned only with variable errors whereas accuracy depends upon freedom from both constant and variable errors. See *constant error, quality, reliable, variable error.*—Monroe, Theory, p. 108f. Symonds, p. 123, 288f.

Achievement age (A. A.) A pupil's age score on an achievement test is usually referred to as his achievement age. A given achievement age, such as 10 years and 8 months or, as it is occasionally expressed, 128 months, means that the pupil who earns this score has done as well on the given test as the average or median pupil whose chronological age is 10 years and 8 months. In actual practice an achievement age is generally established by determining the average or median achievement of a group of pupils whose mental age is the desired amount, in this case 10 years and 8 months. See *age norm, age score.*—Monroe, Theory, p. 155f.

Achievement quotient (A. Q.) This term is applied to a kind of score which shows the relationship between a pupil's actual achievement and what he should achieve. The measure of what he should achieve commonly used is the average or median achieved by pupils of his chronological or mental age. Since, as was explained under achievement age, the average achievement score of a group of pupils of a given mental or chronological age is called an achievement age of the same amount, a pupil's achievement quotient might be secured by dividing his achievement age by either his mental age or his chronological age. The former—that is, division by the mental age—was first suggested and is the common practice, so that usually A. Q. $= \frac{\text{A. A.}}{\text{M. A.}}$

Unfortunately, however, a few persons have introduced confusion by dividing by the chronological age instead of the mental age, so that sometimes A. Q. $= \frac{\text{A. A.}}{\text{C. A.}}$. Since it is the purpose of the achievement quotient to compare a pupil's actual achievement with what he should achieve, it seems distinctly preferable to use his mental age, which is a measure of his ability, as a denominator rather than his chronological age, which merely measures the length of time he has happened to live. See *quotient score.*—Freeman, p. 285f. Kelley, p. 6f., 22f. Monroe, Theory, p. 157f.

Achievement ratio (A. R.). Because the achievement quotient is computed in two ways and hence has two different meanings, it has been proposed that the situation be simplified by restricting it to one meaning and applying the term achievement ratio to the other. Unfortunately there has been no general agreement as to which expression should be called the achievement quotient and which the achievement ratio. It appears, however, that the most frequent use of achievement ratio has been to refer to the result obtained by dividing achievement age by mental age; that is, A. R. $= \frac{\text{A. A.}}{\text{M. A.}}$. Its use in this sense is urged by those who secure the achievement quotient by dividing achievement age by chronological age. See *ratio score.*—Kelley, p. 8. Monroe, DeVoss, and Kelly, p. 381. Otis, p. 172f.

Achievement test. This name is applied to a test which measures a pupil's knowledge or mastery of the subject matter taught in school. In other words, such a test measures what the pupil has learned rather than his capacity to learn.

A. D. Abbreviation for *average deviation,* better called *mean deviation.*

Age norm. An age norm expresses the average or median achievement, intelligence, or other characteristic of a group of pupils of the designated chronological age. In determining age norms for achievement tests, the pupils are frequently grouped according to mental age as this type of grouping is easier to secure than one based on chronological age. Since a given mental age represents the average intelligence of pupils of the same chronological age, the result is the same as if chronological age groups were used. Unless otherwise stated an age norm is usually the average or median of scores made by pupils ranging from the designated age up to the next. For example, a score given as the norm for nine-year-old children is ordi-

narily understood to be for children who are at least nine years of age but not yet ten. See *norm.*—Ruch and Stoddard, p. 346f. Symonds, p. 255f.

Age score. Pupils' scores, both on tests of intelligence and on those of achievement, are frequently expressed in terms of ages, the mental age being used in the case of intelligence and the achievement age in that of achievement. Point scores are transmuted into age scores on the basis of age norms. For example, if a pupil makes a score of 48 upon a particular test and 48 is the age norm for nine years, this pupil is said to have an age score of nine years. An age score of any given amount indicates that the pupil earning it is just at the average of pupils of his age. See *achievement age, age norm, educational age, mental age, social age, subject age.*—Freeman, p. 81f. Monroe, DeVoss, and Kelly, p. 380.

Age variability unit. Among the units employed in educational and psychological measurement is the age variability unit. Such a unit is a function of the variability of a single age group. It is assumed that the variability of a group of pupils of any single age may be equated to that of a group of any other age. Therefore some function of this variability, such as the difference between the average score made by the pupils of an age group and the score dividing the upper 25 per cent from the lower 75 per cent of the same group, is used as a standard unit and considered equal to the same function for a group of any other age.—McCall, How to Measure, p. 272f.

Alternative test. This expression is often applied to one of the chief types of tests included by the new examination and used in many standardized tests. Each item in this type of test permits the pupil a choice between two possibilities, one of which is right and the other wrong. The most common varieties of exercises of this sort are true-false statements and yes-no questions, but others are sometimes used. See *true-false test, yes-no test.*—Odell, Objective Measurement, p. 9f.

A. M. Sometimes used as the abbreviation for *assumed mean.*

Analogies test. Such a test is of the form of the ordinary mathematical proportion, with one of the four terms or occasionally even two of them omitted. An example from the field of algebra is: a^2 is to a^6 as x^3 is to .....; another, from grammar: ran is to run as..... is to sit. This type of exercise is often used in general intelligence tests and sometimes in achievement tests.—Odell, Objective Measurement, p. 27.

Analogy test. Occasionally used as synonymous with *miniature test.*

Aptitude test. Synonymous with *prognostic test.*

A. Q. Abbreviation for *achievement quotient,* also *accomplishment quotient* and *attainment quotient.*

A. R. Abbreviation for *achievement ratio,* also *accomplishment ratio* and *attainment ratio.*

Arithmetic average (Aver. or A.). This is the same as the ordinary *average,* better called the *mean.*

Arithmetic mean (M.). Synonymous with *mean.*

Array. A single row or column of a correlation table including the frequencies which fall in it is called an array. In other words, an array includes all of the measures in a correlation table which fall within a single class or interval of one of the two variables concerned. For example, if age divided into intervals of years is correlated with height by inches, all of the frequencies for each age class, such as 10 years, form an array, as likewise do all for each height class, such as 52 inches. See *correlation table.*

Association test. There is some difference of practice as to the use of this expression. It has been applied to several kinds of tests often included in standardized and new-type tests. Probably its most frequent use has been to designate tests in each exercise of which one, or sometimes more, terms are given to which the pupils are asked to add others closely associated. Sometimes the association is described as *fixed* to designate the fact that the pupil is expected to recognize certain requirements in responding to the exercise; in other cases it is *free.* Thus a list of words may be given for each of which the pupils are to supply a synonym or perhaps an antonym, a list of cities may be given for each of which an important product is to be named, or a list of historical characters for each of whom one important event is to be given.—Odell, Objective Measurement, p. 21f. Russell, p. 124f.

Ass. M. Abbreviation for *assumed mean.*

Assumed average. Synonymous with *assumed mean.*

Assumed mean (Ass. M. or A. M.). In the short method of computing the mean, the standard and mean deviations, and various other statistical expressions, use is made of an assumed or guessed mean. In other words, the person making the calculations inspects the distribution of data and estimates or assumes the value of the mean. This assumed mean is always taken as being the mid-point of a class or interval, and it is almost always desirable that the mid-point selected be as near as possible to the true mean; that is, nearer to it than the mid-point of any other class would be to that mean. If, however, the

guess made is not accurate enough to produce this result, no error will be introduced into any of the succeeding calculations except in the case of the mean deviation.—Odell, Educational Statistics, p. 68f. Rugg, p. 121f.

Assumption. A great deal, if not all, of educational research, especially in the field of measurements, is either explicitly or implicitly based upon assumptions. In some cases these assumptions are apparent facts or principles which cannot be definitely proven, but which appear to be in accord with such evidence as is available. In other cases the asssumptions made are rather of the nature of limitations or perhaps bases for investigation; that is, one may assume that certain things are facts and proceed to investigate or determine what results or conclusions follow. It is probably true that many more assumptions are made implicitly than are definitely stated. In many studies it is, for example, assumed without proof or even without comment that children should attend school, that they should study certain subjects, that they should progress from grade to grade, and so forth.—Monroe, Theory, p. 21f.

Attainment age (A. A.). Sometimes used as synonymous with *achievement age.*

Attainment quotient (A. Q.). Sometimes used as synonymous with *achievement quotient.*

Attainment ratio (A. R.). Sometimes used as synonymous with *achievement ratio.*

Attenuation. If, as is practically always the case, there are chance or variable errors in the measures or scores of either one or both of the two variables involved in a correlation, the effect of these errors is to lower the obtained value of the coefficient of correlation below what it would be if the measures or scores were accurate. This effect—that is, the lowering of the value of the coefficient, is called attenuation. If two series of measures of each of the variables are available, any one of several formulae may be employed to correct for attenuation and give an approximately true value of the coefficient of correlation.—Monroe, Constant and Variable Errors, p. 28f. Odell, Educational Statistics, p. 181f.

Average (Aver. or A.). The term average is employed in two different senses, but to avoid confusion it is better to limit it to one. This is its use as a general term to include the mean, median, mode, geometric mean, and all other measures of central tendency. Its other use is that common in elementary arithmetic and in ordinary conversa-

tion. In this sense it refers to the sum of a number of measures or quantities divided by their number. It is recommended by most statisticians, however, that the term mean be used in this latter sense. See *central tendency, mean.*—Odell, Educational Statistics, p. 64f. Otis, p. 6f. Rugg, p. 99f.

Average deviation (A. D.). Synonymous with *mean deviation.*

b. Abbreviation for the *coefficient of regression.* Subscripts, usually x and y or 1 and 2, are employed to distinguish between the regression coefficients of the two variables concerned in an ordinary regression or correlation.

Battery of tests. A group of several tests, usually achievement tests in several subjects, given pupils as part of a single testing program either at one time or within a short period of time, is frequently called a battery of tests. The term is more or less but not absolutely synonymous with the expression *general survey test.*—Russell, p. 178f.

Best-answer test. Synonymous with *multiple-answer test.*

Best-reason test. This is a variety of the best-answer or multiple-answer test. The suggested answers are reasons rather than mere facts or other items.

Bi-modal. A graph or distribution which has two modes—that is, two points at which the frequencies or numbers of cases are greater than on either side of each, is called bi-modal. In such cases the mode at which the number of cases is the greater is called the major mode; the other, the minor mode. See *mode.*

B-score. This expression is practically synonymous with *grade score.* It consists of one figure in units' place indicating the grade and one in tenths' place indicating the month of the school year, thus assuming a school year of ten months. To illustrate, a B-score of 4.3 is the average for fourth-grade pupils in the third month of the school year. Point scores are transmuted into B-scores by the same general method as into any other derived scores; that is, the average or median point score for each given grade and each month of the school year is determined. The name B-score was proposed in honor of Binet and Buckingham.

C. A. Abbreviation for *chronological age.*

Cause and effect test. This name is applied to a form of test often used as part of a new-type examination, and also sometimes in standardized tests. Each exercise therein consists of several words or phrases one or more of which are causes and the remaining ones, effects. Pupils are instructed to mark all the causes or all the effects

by underlining or by some other method. This form of test is sometimes classed under association tests and also sometimes under multiple-answer tests.

C. B. Abbreviation for *coefficient of brightness.*

Central tendency. The point on the scale about which the measures composing a frequency distribution tend to group themselves is called the central tendency. Any average, using this term in its wider sense, is a measure of central tendency. See *average, mean, median, mode.*—Odell, Educational Statistics, p. 64f. Otis, p. 6f. Rugg, p. 97f.

Chance error. Synonymous with *variable error.*

C. I. Abbreviation for the *coefficient of intelligence.*

Class interval (i). This expression, sometimes shortened to interval, refers to the width of a step, class or group in which measures are grouped in a frequency table. For example, if in tabulating pupils' ages all those from six years up to but not including six years and six months are grouped together, those from six years and six months up to but not including seven years are also grouped together, and so on, the class interval is six months.—Odell, Educational Statistics, p. 17. Rugg, p. 83f.

Classification test. This expression is employed in at least two senses. One usage refers to any test designed primarily for classifying school pupils for purposes of instruction. The second meaning refers to a variety of the new examination. Each exercise in this variety consists of a number of terms several of which are alike in some way. The pupils may be instructed to underline or otherwise indicate the words which are alike or to mark those which are unlike the majority. —Odell, Objective Measurement, p. 26f.

Coefficient of brightness (C. B.). The coefficient of brightness is a rarely used measure of intelligence compared with chronological age, similar to but not identical with the intelligence quotient. Theoretically the two are the same for children up to the age of fourteen years. In the extreme ranges, however, it is unlikely that they will correspond exactly. The coefficient of brightness is obtained by dividing a pupil's score by the score which is normal for his age. This measure has now been displaced by the index of brightness. See *index of brightness.*—Otis, p. 153f.

Coefficient of correlation (r). There are a number of numerical expressions or indices of correlation which may be called coefficients of correlation. The term is, however, generally restricted so that it

applies only to the one obtained by the product-moment method and abbreviated by r, which is the most frequently used measure of correlation. This is sometimes called the Pearson coefficient because its use was strongly advocated by the English statistician, Karl Pearson. It is an index of rectilinear or straight-line correlation or relationship which ranges in value from +1.00 through zero to —1.00. A value of +1.00 indicates perfect positive correlation, one of zero no correlation at all, and —1.00 perfect negative correlation. The basic formula for it is $r = \frac{\Sigma xy}{N\sigma_x\sigma_y}$ or $= \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$. See *correlation, negative correlation, positive correlation.*—Odell, Educational Statistics, p. 150f. Odell, Interpretation, p. 33f. Otis, p. 181f.

Coefficient of correspondence. The coefficient of correspondence may be defined as the per cent of individuals who have the same relative position within the whole group in one series of measures as they do in the other of the two being compared. It will be seen that the meaning of this definition depends upon the interpretation of the words "have the same relative position." Since different statisticians and others have defined "the same relative position" differently, there are a number of ways in which coefficients of correspondence have been computed.—Odell, Educational Statistics, p. 299f.

Coefficient of intelligence (C. I.). In connection with a few intelligence tests it has been recommended that instead of using the intelligence quotient, the ratio of a child's score to the average score of a child of his own age, called the coefficient of intelligence, be employed. As is true in the case of the intelligence quotient, a coefficient of intelligence above 1.00 indicates superior mentality, one of 1.00 exactly normal or average mentality, and one below 1.00 inferior mentality. Because of the difference in methods of computation it cannot be assumed that a coefficient of intelligence of any given amount other than 1.00 means exactly the same as an intelligence quotient of the same amount.—Freeman, p. 134, 281f.

Coefficient of multiple correlation ($R_{1 \cdot 23 \ldots n}$ or $R_{1(23 \ldots n)}$). The coefficient of multiple correlation is a product-moment coefficient derived from ordinary or simple product-moment coefficients of correlation. See *multiple correlation, product-moment correlation.*—Odell, p. 252f. Otis, p. 239f.

Coefficient of partial correlation ($r_{12 \cdot 34 \ldots n}$, $r_{123 \cdot 45 \ldots n}$, etc.). The coefficient of partial correlation is derived from simple product-moment coefficients of correlation and is itself a product-moment coefficient measuring the degree of partial correlation. See *partial correlation, product-moment correlation.*—Odell, p. 245f. Otis, p. 232f.

Coefficient of regression (b). This is an expression which shows the average change in one of two associated variables for each unit change in the other. Thus if the coefficient of regression of one variable on the other is .75 it means that on the average the first variable will increase .75 for every increase of one unit in the other, and will decrease .75 unit for every decrease of one. The formula for the coefficient of regression of one variable, X, on the other, Y, is $b_x = r \frac{\sigma_x}{\sigma_y}$. —Odell, Educational Statistics, p. 189f. Rugg, p. 248f., 254f.

Coefficient of reliability. The coefficient of reliability is merely the coefficient of correlation between the scores secured from two applications of the same test or of duplicate forms thereof. The two applications should be separated by only a short interval of time so that as little change as possible will occur in the intelligence and knowledge of the pupils tested. A coefficient of reliability above .90 is relatively high for a group test. Most of those of the best group tests run from .90 down to perhaps .70. For several individual tests and even two or three of the longest group tests, the coefficients of reliability are above .95. See *coefficient of correlation, reliable.*—Monroe, Theory, p. 202f. Odell, Educational Statistics, p. 185f. Ruch and Stoddard, p. 355f.

Coefficient of validity. This name is given to a coefficient of correlation between test scores and some criterion measure by which the validity of the test is being judged. See *coefficient of correlation, criterion measure, validity.*

Column diagram. Synonymous with *histogram.*

Combined dimensions. Instead of describing each characteristic or dimension of pupils' performances separately, the directions for scoring some test papers provide for a single combined description or measure of two or in some cases three dimensions. For example, if the number of exercises done correctly is taken as the score on a uniform test, this score represents a combination of rate and accuracy. If a scaled test has a time limit short enough that pupils do not reach their limits of difficulty and if the number of exercises done correctly is taken as the score, the result is a combination of all three dimensions, rate, quality, and difficulty. See *dimensions of pupils' performances.*—Monroe, Theory, p. 130.

Comparable measures. Measures are said to be comparable when they are expressed in terms of the same unit and with reference to the same zero point. The ordinary method of rendering the scores on two tests comparable is to change those on one to the scale used on the

other. Sometimes both are changed to a common scale different from that of either. Several different methods of doing so have been recommended.—Monroe, Theory, p. 211f. Odell, Educational Statistics, 295f.

Completion test. One of the most common forms of the new examination is the completion test. Such a test usually consists of a number of statements or sentences in each of which one or sometimes more of the important words have been omitted and are to be filled in by those being tested. Sometimes a completion test takes the form of a connected paragraph. This form of exercise is also employed in many standardized tests.—Odell, Objective Measurement, p. 12f. Ruch and Stoddard, p. 267, 273. Russell, p. 147f.

Composite score. A composite score is the average or mean of the scores yielded by several tests after they have been expressed in terms of a common unit and from a common zero point so that the process of averaging is justified. In other words, the scores must be made comparable before being averaged. If they have not been so expressed the resulting mean is liable to have no significant meaning. The term is often limited to the mean of scores from tests in the same field.—Monroe, Theory, p. 224f. Russell, p. 267f.

Comprehensive examination. A comprehensive examination is one, usually of the new type, which tests knowledge over a wide field of subject matter rather than intensively on a comparatively few topics.

Constant error. A constant error is one which tends to be in the same direction for all members of a given group of pupils. Frequently also it is approximately uniform, either absolutely or relatively, for all the individuals included. The group concerned may be of any size from a portion of a class to all the children in a school system or group of systems. As an example of absolute constant errors, those resulting from measuring the heights of children who stand against the wall with their heels upon the quarter round may be cited. In this case the heights of all would be in error by the same or approximately the same amount. On the other hand, if heights were measured with a foot-rule one-half inch too short, the absolute magnitudes of the errors would depend upon the heights, but their relative size would be approximately the same; that is, about $\frac{1}{24}$ of the height of each individual measured since $\frac{1}{2}$ inch is $\frac{1}{24}$ of a foot. Constant errors do not affect the coefficient of correlation, but do affect the mean and all other measures of central tendency. Any such measure will be in error by an amount equal to the average of the constant errors in the data from which it is derived. See *variable error.*—Monroe, Constant and Variable Errors. Monroe, Theory, p. 198, 243.

Content examination. The term content examination is used to refer to an achievement test or examination over the school subjects as distinguished from an intelligence test or a prognostic test not covering specific subjects already studied.

Control group. In carrying on experimentation in education it is very common to make use of two or more groups of pupils, usually though not necessarily equivalent. If there are only two groups, one of them, and if there are a larger number than two, one or more, are control groups. The pupils in control groups are subjected to the same measurements as those in the other or experimental groups but not to the experimental methods or procedures being tried out. Therefore the results in these groups serve as a basis of comparison for those obtained in the experimental groups and thus supposedly indicate how much of the gain or change produced in the latter group may have resulted from the experimental methods or procedures. See *equivalent groups method.*

Control of testing conditions. One of the most important essentials in the determination of norms or of scores to be compared with norms or other scores is that there be satisfactory control of the testing conditions under which the scores are obtained. These testing conditions include all factors other than pupils' abilities or knowledge which affect or determine their performances. Among the most important of these factors are the explanation of the tests to the pupils, the time allowed for their work, the form in which the tests are presented, the pupils' physical condition and emotional status, and the effort which they put forth. There is said to be satisfactory control of testing conditions when all such factors are made the same for all pupils taking the test or when the amounts of variations occurring in any of the factors are known.—Monroe, Theory, p. 81f.

Correlation. The relationship between two or more series of measures of the same individuals is called correlation. Another definition is that the method of correlation is the study of paired facts. For example, one may wish to compare pupils' marks in arithmetic with their marks in reading; that is, to compare the mark of each pupil in one subject with his mark in the other, or to compare pupils' heights and weights. Such a comparison is usually summarized by statistical methods into a single figure or index. Of such indices the coefficient of correlation is the most commonly used, but the ratio of correlation, and coefficients of rank correlation, of partial correlation, of multiple correlation, and other indices are sometimes employed. If the two series of measures or variables being compared vary together;

that is, if as one increases the other also increases, the correlation is said to be positive or direct; whereas if as one increases the other tends to decrease, it is said to be negative or inverse. The coefficient of correlation and some of the other measures used range in value from +1.00, denoting perfect positive correlation, through zero, denoting no correlation at all, to —1.00, denoting perfect negative correlation. On the other hand, the ratio of correlation and several of the other measures are always positive, ranging from 1.00 down to zero, and thus do not distinguish between positive and negative correlation. It is perhaps worth noting that the existence of correlation does not at all imply causation. To illustrate, if a high correlation is found between pupils' marks in reading and their marks in arithmetic, it is not proof that one causes the other. Both may be caused by a third factor or the connection may be even more indirect than this. See *coefficient of correlation, multiple correlation, partial correlation, rank correlation.*—Odell, Educational Statistics, p. 147f. Otis, p. 175f.

Correlation coefficient (r). See *coefficient of correlation.*

Correlation graph. A correlation graph is in many ways similar to a correlation table. The difference consists in the fact that instead of containing numbers which would show the number of cases in each compartment of the table, it contains dots or other marks which show the location of the various cases on a graph constructed on the X- and Y-axes commonly used in mathematical work. See *correlation table.* —Odell, Educational Statistics, p. 156f.

Correlation ratio (eta, η). See *ratio of correlation.*

Correlation table. A correlation table is a two-way or double-entry table which shows the relationship between two series of measures of the same individuals or, in other words, of a set of paired facts. If more than a small number of cases are concerned in the computation of a coefficient of correlation, the data are almost always put in this form. The scale used in measuring one of the two variables is laid out in a horizontal direction and that of the other vertically. The entry in each square or compartment of the table indicates the number of cases for which one of the measures has the value indicated by the scale value of the row, and the other measure that of the column, in which the entry occurs. For example, suppose that the two variables correlated are age and score on an intelligence test; that ages have been grouped by years on the horizontal scale and test scores by intervals of five points on the vertical scale. If the number 8 occurs in the column headed 9-9-11 and in the line labelled 45-49, it means

that there are eight children of age nine or above but not yet ten who scored from 45 to 49 inclusive on the test.—Kelley, p. 158f. Odell, Educational Statistics, p. 156f.

Criterion. The term criterion is applied to any principle, law, fact, or other standard by which validity may be determined. This includes not merely the validity of a test or scale but also of the selection of cases or items, of a basis of comparison, a statement of a problem, an assumption, a method of procedure, or any other step involved in research.—Monroe, Theory, p. 183f. Monroe and Engelhart, p. 57f. Ruch and Stoddard, p. 45f.

Criterion measure. A criterion measure is any measure which may be used as a basis for comparison or correlation to determine the validity of the scores yielded by a given test. Teachers' estimates of achievement and sometimes of intelligence, school marks, school grade, the composite scores from a number of tests, and sometimes the scores from a single other test, are among the criterion measures most commonly used. It should perhaps be noted that for group tests of intelligence a very common criterion measure has been the Stanford Revision of the Binet-Simon Scale.—Monroe, Theory, p. 221f.

Critical attitude. This attitude requires that assumptions, data, conclusions, and all other activities or procedures be subjected to critical scrutiny to determine their validity for the purposes for which they are employed. To state it differently, the critical attitude requires that an investigator have an unprejudiced attitude and carefully weigh all the evidence at hand before arriving at any conclusion. It also requires that the conclusions reached be considered more or less tentative rather than final and always subject to revision in the light of any fresh evidence which appears to justify revision. See *scientific*.

Cross-out test. This name has been applied to various varieties of the new examination in which pupils are required to cross out certain items. Probably its most frequent application has been to the form of association or multiple-answer test in which several terms are given and the one or perhaps more not connected with a given term or similar to the majority are to be crossed out. It is also used in a number of standardized tests.

Crude data. Data are said to be crude when they are not highly exact or accurate but are merely comparatively rough approximations. This condition is usually due to the use of measuring instruments that have rather large units or are in some other way relatively unrefined. Thus if pupils' heights are measured with a foot-rule containing no

divisions, the resulting measurements are very crude. If heights are measured with a ruler divided into inches but not into fractions of inches the resulting measurements are still somewhat crude.

Crude score. This expression is used in two slightly different ways. In one the adjective crude has the same meaning as in the expression crude data explained just above. In the other crude score may be considered as synonymous with *raw score.*

C-scale. The C-scale is similar to the T-scale, the chief difference being that the unit used is .1 quartile deviation instead of .1 standard deviation. The scale extends the same distance as the T-scale; that is, from five standard deviations below the mean to five above the mean, and therefore since the quartile deviation is only about two-thirds the standard deviation, it is composed of 148 units instead of the 100 of the T-scale. Comparatively few tests provide for the use of the C-scale. See *T-scale.*

C-score. A score given according to the C-scale. The range of such scores is from zero through 74, the average, up to 148. Such a score indicates the point on the scale at which the difficulty is such that the pupil receiving this score can respond correctly to just half the exercises of that difficulty.

Cumulative frequency curve. Synonymous with *ogive.*

Cumulative frequency table. A cumulative frequency table is one in which the frequencies or entries indicate the total number of cases either in and below, or in and above, as the case may be, the given class. The former is most common. Such a table is generally constructed from an ordinary frequency table. To make a cumulative table indicating the total number of cases in and below, the frequencies in an ordinary frequency table are summed up to and including each class to obtain the cumulative frequency for that class. For example, if there are 2 cases in the lowest class, 3 in the next to the lowest and 6 in the next, the cumulative frequency for the latter is 11, found by adding 2, 3 and 6. For a cumulative table showing the number of cases in and above the ordinary, frequencies are summed down to and including each class to yield the cumulative frequency for it.—Odell, Educational Statistics, p. 30f.

Curvilinear relationship. The term curvilinear is used in contrast to rectilinear to apply to cases in which the best graphic representation of the relationship between two variables is a curved rather than a straight line. That line of relationship from which the total deviation or departure of the measures is the least is considered the best fitting

line. If the departure from a straight and a curved line is the same, the former is preferred. The most common, indeed practically the only, expression employed as an index of curvilinear relationship is the ratio of correlation. See *ratio of correlation.*—Odell, Educational Statistics, p. 207f.

Cycle test. A cycle test consists of exercises or items differing in difficulty or perhaps in form or kind, but so arranged that the variations occur in cycles. For example, a cycle of four might be used, in which case the first, fifth, ninth, and so forth exercises would be similar; likewise the second, sixth, tenth, and so forth would be similar; also the third, seventh, eleventh, and so forth; and the fourth, eighth, twelfth, and so forth. A cycle test may be treated as a uniform test as regards both administration and scoring without introducing serious errors. Its use is to be recommended when it is desired to include within a single test exercises of several levels of difficulty or of several different sorts and to make sure that all pupils attempt some of each difficulty or sort.

D. This letter is used as an abbreviation in several different connections. Perhaps the most common of these is that D is used for difference in one method of rank correlation. The difference referred to is that between the rank of a case in one series of measures and its rank in the other. D is also frequently used as an abbreviation for the *10-90 percentile range.* Sometimes D is the abbreviation for *decile,* but *Dec.* is better used in this connection.

Data. The data employed in educational research are not limited to collections of statistical facts, but also include historical facts, principles, opinions, and items of various other sorts.—Monroe and Engelhart, p. 27f. Rugg, p. 28f.

Dec. Abbreviation for *decile.* The subscripts 1, 2, and so on up to 9 are used to indicate the first decile, second decile, and so on up to the ninth.

Decile. The deciles are the points which divide the total number of cases contained in a frequency distribution into ten equal parts; that is, into ten parts each of which contains the same number of cases. Thus one-tenth of all the cases lie at or below the first decile and nine-tenths at or above it, two-tenths at or below the second decile and eight-tenths at or above it, and so forth. Occasionally the term decile is also applied to one of the ten parts mentioned above.—Odell, Educational Statistics, p. 111f.

Definition of problem. To define a problem is to determine and state the particular questions that are to be answered. Some problems involve only one or two questions; others include several. Whatever the number, the formulation in precise terms of each question and subordinate question to be answered is the first step in educational research. If assumptions are made, as is commonly the case, they should be stated. It is also necessary to specify limitations and to define terms that do not have precise meanings or signify the same to all persons.—Monroe and Engelhart, p. 14f.

Derived measure. A derived measure is one which is derived or computed from the original measures obtained. It may be derived by a very short and simple process or it may require a long and complex one. Among the most common derived measures are the mean, the median, the mode, the quartile deviation, the standard deviation, the mean deviation, the probable error, the coefficient of correlation, the ratio of correlation, and the coefficient of regression. Derived measure is also sometimes used as synonymous with *derived score* or *transmuted measure.*

Derived score. Except by chance, two or more tests do not yield point scores expressed in terms of the same unit or from the same zero point. Therefore a number of proposals have been made looking to the calculation and use of scores which describe pupils' performances in terms of a unit and zero point constant for all tests or at least for a large number of tests. Such scores are called derived scores. They include age scores, grade scores, quotient scores, percentile scores, T-scores, and others.—Monroe, DeVoss, and Kelly, p. 380f. Symonds, p. 310f.

Deviation. The spread or scatter of a set of measures about a point, which is almost always a measure of central tendency—that is, an average—is called deviation. It is commonly measured by any one of five or six measures of deviation or variability each of which yields a summary statement from a slightly different standpoint. These measures are the range, the mean deviation, the median deviation, the quartile deviation, the standard deviation, and the 10-90 percentile range.—Odell, Educational Statistics, p. 117f. Rugg, p. 149f.

Diagnostic test. A diagnostic test is one which yields detailed information concerning pupils' achievement in one or perhaps more relatively restricted fields. This type of measuring instrument frequently consists of several sub-tests which yield separate measures of pupils' achievements in a variety of fields. Such a diagnostic test can be used as a survey test by employing some procedure for combining

the scores yielded by the separate sub-tests into a single score. The primary purpose of diagnostic tests is to point out the specific weaknesses of pupils as a basis for remedial instruction.—Monroe, Theory, p. 40.

Difficulty. Difficulty is one of the three characteristics or dimensions of pupils' performances. It has been defined as that characteristic of an exercise which when present in a large degree causes a large per cent of incorrect responses, and when present in a small degree, a small per cent of incorrect responses. In other words, the degree of difficulty of an exercise is determined by the per cent of incorrect responses obtained when it is given to a large number of pupils. If the point of zero difficulty is determined and if certain assumptions are made concerning the distribution of ability of the group of pupils to whom an exercise is given, the degree of difficulty of an exercise can be expressed in terms of a measure of the variability of this distribution of ability. This unit is the difference in difficulty between two exercises each of which is answered correctly by a certain given per cent of pupils, the two given per cents of course being different. The median deviation, usually incorrectly called the probable error, and the standard deviation are the two units most commonly used for this purpose. Thus the difficulty of an exercise may be described as being 1.4 P. E., 2.5 P. E., 1.2 σ, and so forth.—Monroe, Theory, p. 61f.

Difficulty score. A difficulty score is a statement of the highest level of difficulty on which a pupil has responded to the exercises with a specified or standard degree of accuracy. Sometimes 100 per cent accuracy is required, sometimes 50 per cent accuracy, and occasionally some other per cent. Such a score is yielded only by scaled tests. See *difficulty*.—Monroe, Theory, p. 94f., 118f. Russell, p. 226f.

Dimensions of pupils' performances. Pupils' performances are described in terms of three distinguishing characteristics or dimensions. These are (1) the amount, or, when produced under timed conditions, the rate of work, (2) the quality or accuracy of the performance, and (3) the level of difficulty upon which it is given.—Monroe, Theory, p. 19f.

Direct correlation. Synonymous with *positive correlation.*

Directions test. A directions test is one which measures the ability of pupils to carry out directions as given. Such a test is found as a part of a number of intelligence tests.—Freeman, p. 262.

Discrimination. A test is said to possess satisfactory discrimination when the scores earned upon it by pupils who are known to differ in ability vary in accord with these known differences. Thus a test

that is too easy lacks discrimination because a number of pupils make perfect scores, and one that is too hard lacks it because a number of pupils make zero scores. Other evidence also may indicate a lack of discrimination. If a distribution of scores differs conspicuously from the normal distribution when there is reason to believe that the distribution of true scores would approximate the normal, this is evidence that the test does not discriminate satisfactorily among certain pupils. If two groups are known to differ in ability, as for example a fifth-grade group and a sixth-grade group, a test which fails to yield a higher average score for the higher group, in this case the sixth grade, is evidently lacking in discrimination. Furthermore, if the unit used is so large that pupils who differ in ability receive identical scores, the test does not possess satisfactory discrimination. See *undistributed scores*.—Monroe, Theory, p. 219f.

Discussion examination. Synonymous with *traditional examination.*

Dispersion. Synonymous with *deviation.*

Division. As applied to tests, this is usually synonymous with *part.*

Duplicate form. Many standardized tests possess two or more forms usually called Form A, Form B, and so forth, or Form 1, Form 2, and so forth. These forms consist of exercises alike in form and kind, though of course not identical, and are therefore called duplicate forms. In almost all cases such duplicate forms have been constructed with the intention that they shall be of equivalent difficulty, but this result has not always been attained. In its narrower usage the expression duplicate form does not signify such equivalence but as commonly used this is implied. See *equivalent form, form.*—Monroe, Theory, p. 169f. Ruch and Stoddard, p. 65f.

E. A. Abbreviation for *educational age.*

Educational age (E. A.). This expression is almost but not quite synonymous with achievement age. It differs in that it is ordinarily applied only to a pupil's average standing in a number of school subjects expressed in terms of an age score, whereas achievement age may refer to a single subject or the average of several. See *achievement age, subject age.*

Educational guidance. As distinguished from vocational guidance, educational guidance is the advising and directing of pupils in the choice of subjects and other connected matters and not in regard to the choice of a vocation or occupation. The two types of guidance

are, however, closely related and frequently, perhaps usually, must be considered together.

Educational objectives, agreement with. In the selection of exercises or items to be included in a test and of subject matter to be included in a course or curriculum, it is desirable to examine such exercises, items, or subject matter with reference to their agreement with educational objectives. For example, in the construction of his spelling scale, Ayres selected certain words on the basis of their frequency of use in adult correspondence. Charters studied the language errors most commonly made by children and not only incorporated these into his language and grammar tests but also made them the basis of a course of study in this subject. In other cases the consensus of opinion of competent persons, or what amounts to almost the same thing, frequency of occurrence in textbooks, has been employed as a guide in selection.—Monroe, Theory, p. 89f.

Educational quotient (E. Q.). The quotient obtained by dividing a pupil's educational age by his chronological age has been called his educational quotient. That is, $\text{E. Q.} = \frac{\text{E. A.}}{\text{C. A.}}$ Such a quotient shows a pupil's average standing in a number of school subjects as compared with the average of pupils of his chronological age. See *achievement quotient, subject quotient.*—McCall, How to Measure, p. 36f. Monroe, Theory, p. 156f.

Educational ratio (E. R.). Some of those who have advocated that the result obtained by dividing a pupil's educational age by his chronological age be called his educational quotient have also proposed that his educational age divided by his mental age be called his educational ratio. The same result can be obtained by dividing the educational quotient by the intelligence quotient. An educational ratio in this sense is, therefore, synonymous with an achievement quotient in its usual sense if that achievement quotient is the average of quotients in several different subjects. See *achievement quotient.*

Educational research. See *research.*

Educational test. Synonymous with *achievement test.*

Empirical test. The term empirical test is frequently applied to one chosen through the trial and error method. In other words, a number of tests are tried out, usually without any very strong theoretical reason why they, rather than others, should be considered, and the one or ones which appear most useful for the purpose in mind selected. This method of choosing tests has probably received more

use in connection with vocational prognosis or prediction of aptitude than in any other field.

E. Q. Abbreviation for *educational quotient.*

Equivalent form. If the two or more duplicate forms which have been prepared for many standardized tests yield equal or equivalent scores, they are said to be equivalent. In very few, if any, cases is the equivalence perfect, but for many tests it approaches perfection very closely. It is a decided advantage that duplicate forms be equivalent or very nearly so. See *duplicate form, form.*—Monroe, Theory, p. 169f.

Equivalent groups method. This is a method of educational experimentation in which two or more equivalent groups of pupils are used. Different procedures or methods are employed in the two or more groups and the comparison of results at the end of the experiment offers evidence concerning the relative merits of these procedures or methods. In general, groups are considered equivalent when their means and variabilities are the same. It is desirable and for some purposes necessary, however, that the pupils in one group match those in another, taken pair by pair.—McCall, How to Experiment, p. 18, 29f., 40, 161f.

E. R. Abbreviation for *educational ratio.*

Error. There are a number of kinds of errors present in educational data. In most instances their magnitude and number can be determined approximately, but not for any particular individual. See *constant error, error of estimate, error of measurement, error of sampling, variable error.*

Error of estimate. Errors of estimate are those errors involved in estimating the values of one variable from those of another by the use of the regression equation. For example, if the scores of a number of pupils upon an intelligence test and their average school marks have been correlated and the regression obtained, the differences between the estimates of school marks based upon intelligence test scores and the marks actually assigned are errors of estimate. Also if school marks are known and intelligence test scores estimated from them, the differences between estimated and actual scores are errors of estimate. Such errors are usually measured by the standard or probable error of estimate.—Monroe, Theory, p. 199f., 350f. Odell, Educational Statistics, p. 230f. Odell, Interpretation, p. 28f., 41f.

Error of measurement. Errors of measurement are similar to errors of estimate, but differ in that whereas the latter are involved in

estimating one actual or obtained score from another, errors of measurement are those involved in estimating true scores from a series of actual scores. For example, if two equivalent forms of a reading test have been given, the errors involved in estimating Form 2 scores from Form 1 scores, or vice versa, are errors of estimate whereas those involved in estimating true scores from either Form 1 or Form 2 scores are errors of measurement.—Monroe, Theory, p. 207f., 354f. Odell, Educational Statistics, p. 230f. Odell, Interpretation, p. 28f., 41f.

Error of sampling. Errors of sampling occur in derived measures and are due to the fact that such measures are frequently calculated from a limited number of cases chosen as being representative of a larger group or population. In many cases it is either impossible or impracticable to utilize all cases of the sort being dealt with. For example, if one desires to make a study of ten-year-old boys he must do so by using a selected sample of boys of that age, and derived measures computed from this sample contain errors of sampling. If, as is generally assumed, the sample is chosen without bias, the errors in the derived measures will be smaller the larger the sample. Their magnitude decreases in inverse ratio to the square root of the number of cases, therefore since 200 is four times 50 and the square root of four is two, the average magnitude of the errors present in derived measures obtained from a sample of 200 individuals would be only one-half as great as in those obtained from 50 individuals. Errors of sampling are commonly described by stating the probable or the standard error of the derived measure in question. See *random sample, sampling.*—Monroe, Theory, p. 330. Odell, Educational Statistics, p. 221f. Odell, Interpretation, p. 21f.

Essay examination. Synonymous with *traditional examination.*

Eta (η). Abbreviation for the *ratio of correlation.*

Exercise. An exercise is a structural unit of a test, in other words, a unit governed by a single set of directions. Some of the simpler types of exercises merely call for a word to be spelled, an arithmetical example to be worked, or a question to be answered. Others are more complex. Some consist of a number of items. A test usually consists of at least several exercises, but occasionally of a single long one.—Monroe, Theory, p. 56f., 89f.

Experimental coefficient. It has been suggested that instead of comparing the difference between two means or other derived measures directly with the probable or standard error of the difference in order to determine its reliability, a formula yielding what

is known as the experimental coefficient be used for this purpose. This formula requires merely that the difference be divided by 2.78 times the standard error of the difference. In other words, Exp. Coef. $= \frac{\text{Diff.}}{2.78\,\sigma_{\text{diff.}}}$. The resulting experimental coefficient is interpreted by means of a table of chances which shows how likely it is that the difference in question is significant. The smaller the experimental coefficient, the smaller are the chances that it is so. An experimental coefficient of 1.0 is generally accepted as practical certainty.—McCall, How to Measure, p. 404f. Odell, Educational Statistics, p. 228.

Experimental factor. The factor or element in the situation with which one is experimenting is sometimes called the experimental factor. Sometimes only one such factor is involved, sometimes more than one. —McCall, How to Experiment, p. 81f.

Experimental group. One of the most common methods of educational experimentation involves the use of two or more groups of pupils. The one or more of these in which the experimental procedures or methods are employed are generally called experimental groups in contrast with the others which merely serve for checking results and are called control or check groups. It is usually desirable that the experimental and the control groups be equivalent, but often satisfactory if they are not provided the differences between them are known and measured. See *equivalent groups method.*

Experimentation. Although experimentation is only one of the methods of educational research, it has probably received the major part of the attention and emphasis in this general field within recent years. It may be defined as that method which tests theory by a process of trying it out and evaluating the results obtained. Its purpose is to evaluate some one or more of the factors which enter into the educational process. Experimentation should begin with the definition of a problem followed by the setting up of conditions and the carrying out of procedures which contribute to the solution of the problem. The experimenter should maintain and apply the critical or scientific attitude. It has been said that experimentation is the third stage, or perhaps better the third step, in the determination of truth, the first being authority and the second speculation.—McCall, How to Experiment, p. 1f.

f. Abbreviation for *frequency.*

Fact-finding study. A fact-finding study is one in which the chief purpose is to determine and collect facts. Although such studies

are important and necessary, they cannot be said to be complete educational research. In order that the investigation be so classified the facts found must be satisfactorily interpreted and applied.

First quartile (Q.). The first quartile is that point on the scale of measurement used in connection with any distribution or series of measurements at or below which one-fourth and at or above which three-fourths of the measures fall. $Q_1 = l + \frac{\frac{N}{4} - S}{f}$. See *quartile*.—Odell, Educational Statistics, p. 111f.

Foot-rule correlation (R.). One of the two common methods of securing rank correlation is known as the foot-rule method because of the comparative ease with which it may be applied. In the foot-rule formula, which originated with Spearman, the symbol for correlation is R, and the value of R is determined by the differences between the ranks of the measures in the corresponding pairs.—Odell, Educational Statistics, p. 202f.

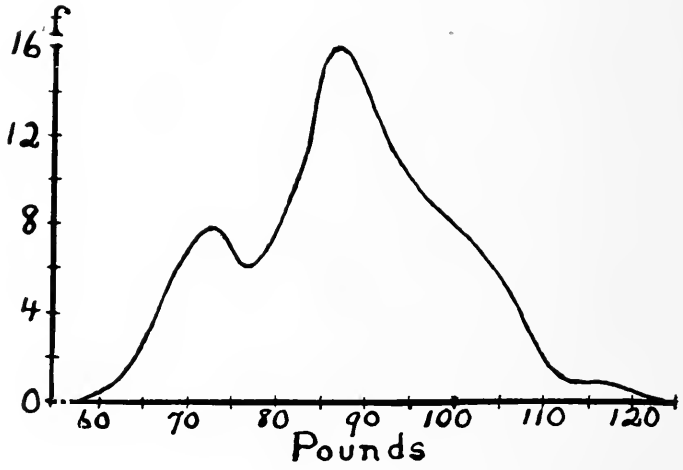
Fore exercise. A fore exercise is a preliminary or trial test which has for its purpose acquainting the pupils with the character of the exercises which they are asked to do in the real test. In administering a test the person doing so should usually see to it that the pupils make the correct responses on the fore exercises. The pupils' performances thereon are not included in computing their scores.

Form. The term form has come to be generally used in the sense of duplicate form. Thus a test is said to have two or more forms when it has two or more measuring instruments consisting of similar but not identical exercises. In a very few cases the word form has been used as synonymous with part, division, or even test. That is to say that Form 1 might be used to indicate the portion of the test for the lower grades, Form 2 that for the upper grades. This usage is, however, so rare as to be practically negligible.

Frequency. The term frequency as a noun is used to refer to the number of measures or cases in a class, or in other words, to an entry in a frequency or correlation table. For example, if in a table of children's weights by five-pound intervals, there are nine cases of children with weights from 75 up to but not including 80 pounds, the frequency in this class is said to be nine. As an adjective frequency is used in a number of connections generally implying that the noun which it modifies refers to a table, graph, or so forth, containing a

number of frequencies. See *frequency curve, frequency polygon, frequency table.*

Frequency curve. This expression is used in two senses, one more inclusive than the other. In the wider sense a frequency curve is any sort of curve or graph which represents a distribution of measures. The three common varieties thereof are the smooth frequency curve, the histogram, and the frequency polygon. All of these are commonly

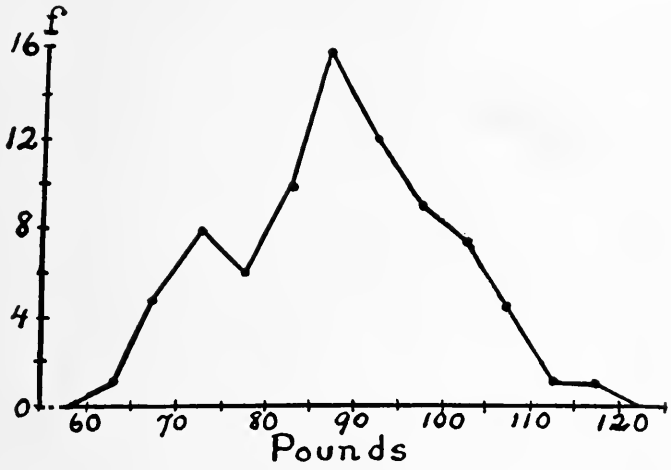


drawn so that the scale of measurement by which the cases included are measured is laid out horizontally, and the scale showing the number of cases or frequencies, vertically. In its narrower sense it refers to a smooth curve which represents a distribution of measures. It is drawn by constructing a smooth curve through points located as for a frequency polygon. A curve of this sort is illustrated by the accompanying figure which represents the distribution of weights of a group of children. It shows, for example, that one pupil of the group had a weight between 60 and 65 pounds, five between 65 and 70, and so on. The greatest height of the curve is above the 85 to 90 interval and shows that more children had weights between these limits than within any other five-pound interval. Also see *normal frequency curve.*—Odell, Educational Statistics, p. 36f. Rugg, p. 88f.

Frequency distribution. Synonymous with *frequency table.*

Frequency polygon. A frequency polygon is one of the three common types of graphs used to represent a distribution of measures.

Its form is illustrated by the accompanying figure. It is constructed by determining and connecting with straight lines a series of points each one of which is directly above the midpoint of a class interval, and at a height equal to the frequency in the class. These points are shown in the figure, which represents the same data as were used for the smooth frequency curve above. See *frequency curve*.—Odell, Educational Statistics, p. 39f. Rugg, p. 90f.



Frequency table. A frequency table consists of one column which indicates the limits of the various classes into which the individual cases included have been grouped and a second which shows the number or frequency of cases in each class. Such a table is illustrated by the columns at the right. The first of these columns designates the various class intervals and the second gives the frequency or number of cases in each. In this example the class intervals are designated in the most common way; that is, by giving only the lower limit of each class. It is then understood that a given class includes all measures from the given lower limit up to the lower limit of the next class. For example, the first class in the table—that is, the one at the bottom—includes all cases having magnitudes of from zero up to but not including five; the next one all those from five up to but not including ten, and so on. The figures in the second column show that the frequency in the 0-up-to-but-not-including-5 class is one, that in the 5-up-

	f
30–	2
25–	4
20–	6
15–	9
10–	7
5–	3
0–	1
N =	32

to-but-not-including-10 class, three, and so forth.—Odell, Educational Statistics, p. 16f. Rugg, p. 81f.

Frequency tabulation. Synonymous with *frequency table.*

Function. As used in the field of education function may be considered as synonymous with purpose or aim. The term is most often employed in connection with standardized tests. The function of such a test is described by a statement of the ability which it is intended to measure plus a statement of the type of information concerning this ability which it will yield. A statement of the function of a test should include as specific information as possible concerning what characteristics or dimensions or combination thereof are measured and also some specifications as to its scope, whether general, diagnostic, or prognostic.—Monroe, Theory, p. 18f.

Functional relation. A functional relation is said to exist between two variables if a change in one produces a corresponding proportional change in the other. The relation between the two variables may be very simple, or it may be decidedly complex and require a considerable amount of computation to determine one from the other. The former, a very simple functional relation, may be illustrated by such an equation as $x = 6y$, which merely means that any change in y produces a corresponding change six times as great in x. A more complex functional relation is indicated by such an equation as $x = \sqrt[3]{\frac{y}{2}}$. This equation signifies that as y is changed in a given ratio, x changes correspondingly according to the cube root of that ratio divided by two. One of the primary assumptions in much if not all educational measurement is that pupils' performances sustain a constant functional relation to the abilities which are being measured.—Monroe, Theory, p. 22, 24.

G. Sometimes used as abbreviation for *geometric mean.*

g. Abbreviation for *gain* in connection with one method of computing *rank correlation.*

G. A. Abbreviation for *guessed average,* better known as *assumed mean.*

General intelligence test. Tests which are designed to measure general intellectual capacity are usually called general intelligence tests in contrast with those designed to measure actual ability in some school subject, which are called achievement tests. General intellectual capacity may be defined as that mental capacity which supposedly may be applied in any field of intellectual endeavor. It has been appropriately

suggested that a more satisfactory name for tests of this capacity would be mental alertness tests, but this term has not come into general use. A majority of the so-called general intelligence tests appear to measure what may be called abstract intelligence as opposed to social and motor intelligence. Most general intelligence tests consist of several sub-tests each of which contains exercises of a particular type designed to test some one manifestation of intelligence. It is assumed that the average or combined score from a number of such manifestations yields a fairly accurate measure of general intelligence.—Freeman, p. 476f. Kelley, p. 4, 116f. Monroe, DeVoss, and Kelly, p. 332f.

General survey test. A general survey test is usually composed of a number of tests or sub-tests each of which covers a different school subject or field of subject matter. Occasionally, however, the term is applied to a test in a single school subject which contains a number of parts covering different phases of the subject. The function of such a test is to yield a general or average measure of pupils' achievements over a comparatively wide field. Ordinarily the scores yielded by the different portions of a general survey test are combined into a single score. Such scores are valuable for determining the general efficiency of a school or teacher, but are rarely of much help in diagnostic and individual work.—Monroe, DeVoss, and Kelly, p. 377f. Ruch and Stoddard, p. 200f.

Geometric mean (G. M., G., or M). This mean is used in dealing with rates of increase. It is the nth root of the product of n measures and therefore must usually be found by the use of logarithms.—Odell, Educational Statistics, p. 94f. Rugg, p. 132f.

G. M. Abbreviation for *geometric mean,* also sometimes for *guessed mean,* more commonly known as *assumed mean.*

Grade. This term is commonly used in two distinct senses. One of these is in such expressions as first grade, second grade, seventh grade, and so forth, to refer to the various stages of advancement in school or units of school organization. The term is also frequently employed to refer to ratings given pupils in such expressions as a grade of 85 per cent or a grade of B. It is decidedly preferable, however, to use the word mark in this second sense and to limit grade to the first meaning given, thus avoiding possible confusion resulting from its double use.

Grade norm. A grade norm is a statement of the achievement or sometimes capacity of pupils in a particular grade. The average or median score of a large number of pupils in a single grade is usually taken as the norm for that grade though rarely some other point is

used. Grade norms are ordinarily based upon the supposition that a school system contains eight elementary grades and four years of high-school work; therefore, if used for comparative purposes in connection with a system which has a different organization, adjustments are necessary. There is no uniformity as to the time of year for which grade norms are given so that this fact should always be stated. See *B-scale, norm.*—Freeman, p. 294f. Monroe, Theory, p. 161f. Ruch and Stoddard, p. 344f.

Grade score. See *B-score.*

Grouped series. Synonymous with *frequency table.*

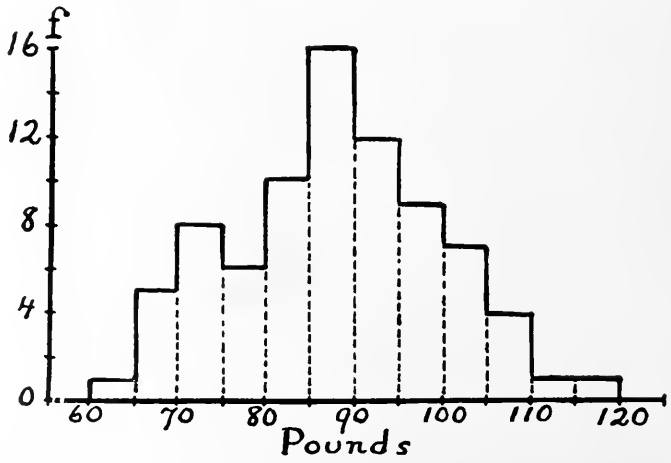
Grouping. This term refers to the classifying or collecting of single measures into classes or groups so that instead of a simple or ungrouped series, a frequency table is formed.—Odell, Educational Statistics, p. 21f.

Group test. A test which can be given to a number of individuals at the same time and by the same examiner is called a group test. Almost all standardized tests are group tests, the chief exceptions being those in oral reading and a few individual ones in intelligence.—Freeman, p. 164f.

Guessed average (G. A.). Synonymous with *assumed mean,* which is a better term.

Guessed mean (G. M.). Synonymous with *assumed mean.*

Histogram. A histogram or column diagram is one of the three common types of frequency curves. It may be thought of as composed of a series of rectangles one of which is erected above each class



interval. The width of each rectangle represents the width of the class interval and its height the number of cases or frequencies in the class. Usually the dividing lines between the rectangles are not shown. The accompanying figure illustrates a histogram with the dividing lines just referred to broken whereas the outside bounding line is solid. The data represented are the same as have already been employed in connection with the smooth frequency curve and the frequency polygon. See *frequency curve.*—Odell, Educational Statistics, p. 41f. Otis, p. 31f. Rugg, p. 91f.

i. Abbreviation for *class interval.*

I. B. Abbreviation for *index of brightness.*

Index of brightness (I. B.). The index of brightness is a measure of intelligence as compared with age. Thus it is in some ways similar to the intelligence quotient or coefficient of intelligence, but it is based upon a fundamentally different assumption. It was suggested by Otis in connection with his general intelligence scales and has not received extensive use in other connections. It is found by calculating the difference between an individual's score and the norm for his age and then according as this difference is plus or minus, adding it to or subtracting it from 100. Thus an index of brightness of 100 is the same as an intelligence quotient of 100, but for other values the two measures are not likely to correspond exactly or even closely.—Freeman, p. 283f. Otis, p. 155f.

Index of reliability. Just as the coefficient of reliability is a measure of the correlation or agreement between the scores resulting from two administrations of the same test or two duplicate forms thereof, so the index of reliability is a measure of the correlation or agreement between one of these sets of actually obtained scores and the corresponding true scores. If the coefficient of reliability is known, the index of reliability is very easily obtained since it is merely the square root of the coefficient. See *coefficient of reliability, reliable.*—Monroe, Theory, p. 206f. Odell, Educational Statistics, p. 188f.

Individual differences. This expression refers to the differences between individuals, usually school pupils, in native ability or capacity, acquired ability or achievement, industry, attitude, interests, health, and the many other characteristics in which they may differ. The frequent occurrence of the term in recent educational and psychological literature and discussions has been due to the fact that until a relatively recent date comparatively few persons realized the number or extent of such differences.—Freeman, p. 367f.

Individual test. An individual test is one which can be administered to only one person at a time. The usual reason is that the subject's responses are oral or that the examiner must note down a rather careful description of them. Except in oral reading there are very few individual achievement tests, but in the field of intelligence testing their use is more common.

Informal test. A test prepared by a classroom teacher is sometimes called an informal test to distinguish it from a standardized test.

Intelligence quotient (I. Q.). The intelligence quotient is by far the most commonly used means of comparing intelligence as measured by a general intelligence test with age. It is found by dividing an individual's mental age, derived from his score on a general intelligence test, by his chronological age. That is, $\text{I. Q.} = \frac{\text{M. A.}}{\text{C. A.}}$. In writing it the decimal point is ordinarily omitted. Thus a pupil whose mental age is the same as the average for all persons of his chronological age, has an intelligence quotient of 100. If his mental age is greater than his chronological age, his intelligence quotient is proportionately greater and if less it is less. For adults and persons in their upper teens the actual chronological age is not used as a divisor, but instead a fixed age supposed to represent the point at which the growth of intelligence ceases is employed. Sixteen has been most commonly used for this purpose though several other ages within two or three years of this have been suggested.—Freeman, p. 98, 276f.

Intelligence test. Synonymous with *general intelligence test.*

Interval (i). Synonymous with *class interval.*

Inventory test. An inventory test is one whose purpose may be said to be the same as that of an inventory or stock-taking in a business establishment. In other words, it is to determine the ability and knowledge of pupils in a certain field at the beginning of a more or less definite period of instruction so that those in charge of the instruction will know the basis upon which they can proceed. An inventory test, therefore, usually covers a particular field of subject matter rather thoroughly. It is more or less synonymous with diagnostic test, but not absolutely so.

Inverse correlation. Synonymous with *negative correlation.*

I. Q. Abbreviation for *intelligence quotient.*

Irregular test. An irregular test is one in which the exercises vary in difficulty and are not arranged in order of increasing or decreasing difficulty. Most tests which contain exercises not selected on

the basis of difficulty are of this sort. In scoring, irregular tests are usually treated as uniform; that is, each item or exercise counts the same amount. Unless the irregularities are extreme, this procedure is unlikely to introduce serious errors in the pupils' scores.—Monroe, Theory, p. 62, 75, 108.

Item. An item is the smallest unit of test construction. Sometimes an item is the same as an exercise; sometimes there are a number of items in a single exercise. Each statement in a true-false test, each blank to be filled in a completion test, each one of several suggested answers in a multiple-choice test, is an item.

Law of the single variable. The law of the single variable is that in making educational measurements, all of the factors which control or affect pupils' performances should be held constant save one, and this one measured. For example, if one wishes to measure rate of reading, such other factors as difficulty of the material read, quality or accuracy of reading, and all the conditions under which the test is given should be controlled or made uniform. A somewhat broader interpretation sometimes given the law of the single variable is that it merely demands the explicit recognition and separate description of the different dimensions, ordinarily three, of pupil performance. Since in many cases it is practically impossible to insure that all the variables except one are constant, this latter interpretation is the one most generally given. See *dimensions of pupils' performances, variable.*—Monroe, Theory, p. 87f.

Lower quartile (Q_1). Synonymous with *first quartile.*

M. Abbreviation for *mean.*

M. A. Abbreviation for *mental age.*

Mark. The term mark rather than grade is best applied to ratings given pupils in terms of per cents, letters, or other symbols. Thus 75 per cent, 88 per cent, A, F, and so on, when used for this purpose are best called marks. By so doing the term grade is restricted to its general use to indicate stage of advancement within a school, such as first grade, fourth grade, and so forth, and thus confusion is avoided.—Symonds, p. 408f.

Matching test. This is one of the forms used in the new examination and standardized tests. In such a test there are two columns of words or other expressions and the pupils are asked to match those in one column with those in the other. For example, the first column may consist of a list of dates, the second of the events which occurred on those dates; the first may consist of a list of Latin words and the

second of their English equivalents, and so forth. It goes without saying that the order of arrangement in the two columns must be different.—Odell, Objective Measurement, p. 18f. Ruch and Stoddard, p. 268f., 276f. Russell, p. 91f.

Md. The most common abbreviation for *median.*

M. D. Abbreviation for *mean deviation.*

Md. D. Abbreviation for *median deviation.*

Mean (M.). The mean is the same measure or quantity as that ordinarily called the average or the arithmetic average in common speech. It is found by dividing the sum of a number of scores or measures by their number. That is to say, $M_x = \frac{\Sigma X}{N}$. The term mean rather than average is preferable in this connection so that the latter can be saved for a more inclusive use and thus confusion avoided. See *average.*—Odell, Educational Statistics, p. 66f. Otis, p. 6f., 17f., 37f. Rugg, p. 114f.

Mean deviation (M. D.). As its name implies, this is the mean or average of the deviations of a set of measures from a given point. Theoretically this point may be any measure of central tendency—that is, any average, using the term in its broad sense; but as a matter of practice the mean deviation is always found around either the mean or the median. For a normal distribution about 57.5 per cent of the scores will not differ from the mean or median by more than one mean deviation and of course the remaining 42.5 per cent will differ by that amount or more.—Odell, Educational Statistics, p. 123f. Rugg, p. 159f.

Med. This abbreviation is sometimes used for *median.*

Median (Md. or Med.). The median is that point on the scale which divides the total number of measures or cases into two equal groups. Thus if there are 80 cases the median is a point such that 40 of the cases lie at or below it and 40 at or above it. Sometimes a distinction is made between a grouped distribution or frequency table and a simple or ungrouped series in that the term median is used in connection with the former and mid-score or mid-measure with the latter. Although such a distinction seems desirable it is not common, but the term median is generally used to include both cases. The formula for the median is $\text{Md.} = 1 + \frac{\frac{N}{2} - S}{f}$.—Odell, Educational Statistics, p. 75f. Otis, p. 11f., 43f. Rugg, p. 103f.

Median deviation (Md. D.). The median deviation is merely the median of the deviations about a given point. The point taken for this

purpose is almost always the mean. Fifty per cent of the scores or measures in a normal distribution lie not more than one median deviation from the mean and the other 50 per cent not less than this distance from it. Although the median deviation could be found by tabulating the actual deviations and determining their median, this method is rarely, if ever, used. Instead the standard deviation is computed and multiplied by .6745 to determine the median deviation. This relationship holds exactly only in case of a normal distribution, but for distributions not extremely different from the normal it is accurate enough for most purposes. The median deviation is often miscalled the probable error, a term which is correctly applied only when it is used in connection with errors. See *deviation, probable error.*—Odell, Educational Statistics, p. 138f. Odell, Interpretation, p. 9f.

Mental age (M. A.). A pupil's score on a general intelligence test expressed in terms of age is called his mental age. To say that a pupil has a mental age of a certain amount—for example, nine years and ten months—means that his intelligence test score is the average or median score made by an unselected or random group of pupils nine years and ten months of age chronologically.—Freeman, p. 84f.

Mental index (M. I.). The mental index is one of the measures of native ability which has been suggested but has received little use. It is determined according to a scale based upon an assumption of normal distribution of ability and such that the lowest possible value is zero, the average or normal value 50 and the highest possible 100. The mental index is, therefore, intended to perform the same function as the intelligence quotient; that is, to compare the intelligence of an individual with the average intelligence of individuals of his age. The method of computing it, however, is distinctly different from that for the intelligence quotient and therefore these two measures cannot be compared directly.

M_g. Sometimes used as abbreviation for *geometric mean.*

M. I. Abbreviation for *mental index.*

Mid-measure. Synonymous with *mid-score.*

Mid-score. The mid-score may be defined as the middle measure of a series of measures or scores arranged in order of size. If there is an odd number of cases it is always an actual measure, but if the number is even the average of the two mid-most measures is taken. This may or may not be the same as any actual measure. For example, the fourteenth of 27 measures arranged in order of size is the mid-score since there are 13 on each side of it. For 28 measures, however, the mid-score must be found by averaging the fourteenth and the fifteenth. —Odell, Educational Statistics, p. 87f. Rugg, p. 109f.

Miniature test. This type of test, which is rarely used except in connection with vocational prognosis, involves a small-scale reproduction of the actual performances in which ability is to be tested. A well-known example of the miniature test was constructed by Münsterberg to predict the ability of motormen. He constructed in the laboratory a chart which represented a street with the various factors and difficulties which must be dealt with in operating a street-car represented upon it. The prospective motormen were required to respond to this situation.—Freeman, p. 412.

Mixed-relations test. Synonymous with *analogies test.*

Mode (Z). The mode of a distribution is that point on the scale at which there are more measures than are to be found at any other point. Thus in a sense the mode may be said to be the typical value or case. In a grouped distribution or frequency table the true mode cannot be determined by inspection but requires rather difficult computation. In such cases it is frequently the practice not to state the mode as a definite point but merely to say that it lies within the interval which contains the greatest frequency. Sometimes one of two or three fairly easy formulae which give approximations to the true mode is employed. The most commonly used of these is that the mode equals three times the median less twice the mean, or $Z = 3Md. - 2M.$ Occasionally the term mode is used in a broader sense to apply to any point on the scale at which the frequency is greater than are the frequencies immediately above and below that point. In this sense a distribution or curve may have two or more modes. In such cases the one at which the frequency is greatest is called the major mode.—Odell, Educational Statistics, p. 89f. Rugg, p. 100f.

M-scale. The M-scale is similar to the much better known T-scale except that it is based upon the ability of a particular group of children and can be used only with that group whereas the T-scale is based upon the ability of twelve-year-old children in general. Both are based upon the assumption of normal distribution of ability and provide scales in terms of which the difficulty of exercises and pupils' scores may be expressed. See *T-scale.*—Russell, p. 269f.

M-score. A score given according to the M-scale.

Multi-modal. A frequency distribution or curve is said to be multi-modal when it includes two or more points at each of which the frequencies are greater than those next to them in each case. In other words, a distribution or curve having more than one mode in the broader sense of the word is called multi-modal. See *mode.*—Russell, p. 221f.

Multiple-answer test. A multiple-answer test is composed of exercises which require pupils to select one or more correct answers out of a group of several given in the exercises. There are many possible forms and varieties of such exercises.—Odell, Objective Measurement, p. 13f. Ruch and Stoddard, p. 267f., 273f. Russell, p. 105f.

Multiple-choice test. Synonymous with *multiple-answer test.*

Multiple correlation. Multiple correlation is the correlation of one variable with two or more other variables in combination. It is almost always expressed in terms of a coefficient of correlation which is computed from the ordinary or product-moment coefficients of correlation between the various pairs of variables involved. See *coefficient of multiple correlation, correlation.*—Odell, Educational Statistics, p. 252f. Otis, p. 238f.

N. This symbol is used as the abbreviation for the total number of cases in a frequency table or any other single group. In cases in which a whole group and a sub-group are dealt with N is commonly used for the entire group and n for the sub-group.

Negative correlation. Correlation or relationship which is such that the larger values of one variable or series of facts tend to be associated with the smaller values of the other and vice versa is called negative. See *correlation, positive correlation.*

New examination. This term has been very commonly employed to include those types of tests or exercises which call for very brief pupil responses in the form of checks, underlinings, single words, and so forth, and which permit objective or near-objective scoring. Among the most common types of exercises included under this heading are multiple-answer, true-false, completion, matching, recall, and analogies. —Odell, Objective Measurement. Ruch and Stoddard, p. 266f. Russell, p. 28f.

New-type examination. Synonymous with *new examination.*

Non-language test. Synonymous with *non-verbal test.*

Non-verbal test. Strictly speaking a non-verbal test is one in which there is no use of words either by the examiner in giving the test or by the subjects in responding to it. Ordinarily, however, the term is more broadly applied to include all tests to which the subjects respond without using language and in which no written directions are employed, regardless of whether or not oral directions are given by the examiner. Such tests are commonly used in testing small children, illiterates, and foreigners.—Freeman, p. 167f., 261f.

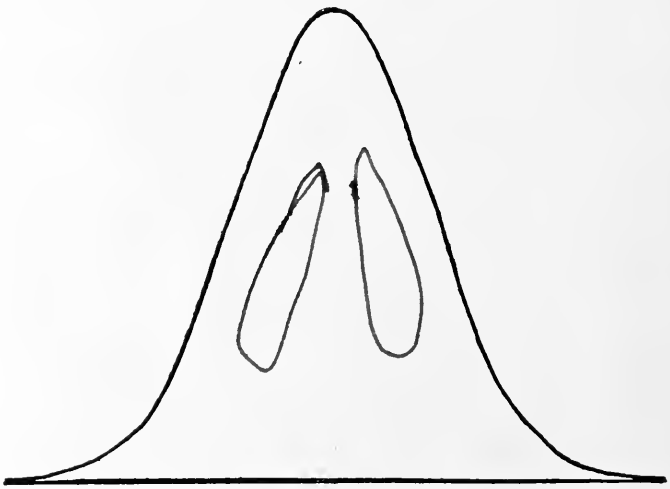
Norm. A norm for a test is a statement of the actual achievement of pupils of the given age or other homogeneous group for which

the norm is being determined. Therefore, a norm is merely a statement of present achievement and not of what achievement should be. It has, however, frequently been used in the latter sense. It is decidedly preferable not to do so but to use the word standard instead whenever reference is made to what pupils should do. In most cases the average or median achievement of a group is taken as the norm, but sometimes other points, such as quartiles or percentiles, are used. Most norms are general norms; that is, they are based upon the scores from fairly large numbers of pupils who are more or less widely scattered over the country. In addition to these, however, local norms for particular states, cities, or even buildings are sometimes used.—Monroe, Theory, p. 161f. Ruch and Stoddard, p. 60f., 343f. Symonds, p. 254f., 265f.

Normal distribution. Synonymous with *normal frequency distribution.*

Normal frequency curve. See *normal frequency distribution.*

Normal frequency distribution. A normal frequency distribution is one which when graphed forms the familiar bell-shaped, symmetrical curve known as the normal frequency curve, the curve of error, the normal probability curve, or the Gaussian curve. As is shown by the accompanying figure, this curve is high in the center, decreases in height rather rapidly near the center, and then more slowly near the extremes. It never actually touches the baseline. The normal distribution occurs more often than any other in educational and other biological data as well as in the operation of the laws of chance when the chances are equal.—Odell, Educational Statistics, p. 52f. Otis, p. 68f. Rugg, p. 191f.



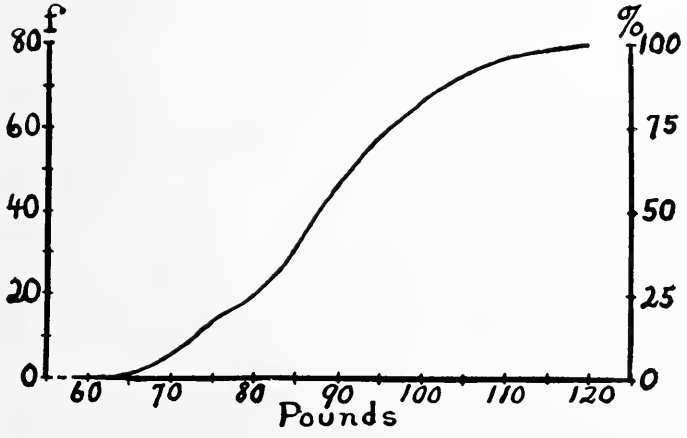
Normal probability curve. Synonymous with *normal frequency curve.*

Objective. This term has two common uses in educational literature, one of which is as a noun and general, the other as an adjective limited to the field of measurement. In its general use objective is synonymous with goal, aim, or purpose, and is frequently used in such phrases as "objectives of education" and "objectives of instruction." According to the second use, a measuring instrument is said to be objective when different persons using it to measure the same thing secure the same results. In other words, a test is objective when there is no doubt in the opinions of competent scorers as to what the correct answers are and when all possible answers must be either definitely right or wrong. In ordinary usage tests which are not absolutely objective, but only approximately or relatively so, are spoken of as objective.—Monroe, Theory, p. 26f., 196f. Ruch and Stoddard, p. 58f.

Objective test. Sometimes the term objective test is used synonymously with new examination, because most of the forms included under that term possess relatively high objectivity. On other occasions it is employed to refer to any test, whether standardized or not, which meets the requirements defined under the second given meaning of objective; that is, which permits no reasonable doubt as to the correctness or incorrectness of all possible answers. See *objective.*

Objectivity. See *objective.*

Ogive. The ogive or cumulative frequency curve is the curve which represents a cumulative frequency table or distribution. It is commonly drawn as in the figure below so that the height of the curve at any given point indicates the total number of frequencies up



to and at that point on the scale of measurement. Sometimes, however, it is drawn in just the opposite manner so that the height at a given point indicates the number of measures at and above that point. The ogive is ordinarily drawn as a smooth curve, though rarely the polygon or histogram form is used. In connection with an ogive it is very common to have two vertical scales. In such cases one of these indicates the actual frequencies and the other the percentile points. In the accompanying figure the column to the left running from zero up to 80 indicates the actual frequencies or numbers of cases and that at the right, running from zero to 100, the percentile points.—Odell, Educational Statistics, p. 49f. Otis, p. 32f., 43f., 53f., 77f.

Omnibus test. An omnibus test is one in which various kinds of tasks or exercises are mixed together in either regular or irregular order instead of being grouped in sub-tests each of which contains exercises of only a single type. Thus there may be an analogies exercise, an example in arithmetic, a statement to be marked true or false, a multiple-answer exercise, a second analogies exercise, a completion statement, and so on. When the term omnibus test is applied in the field of school achievement it is commonly understood that the test covers several different fields of subject matter. This is, however, not necessarily implied by the name.

One-group method. This is a method of experimentation in which an experimental procedure is tried out with a single group and the results which occur in that group noted.—McCall, How to Experiment, p. 14f.

Opposites test. This form of test is one of the new examination types and is also used in some standardized tests, especially those of intelligence and vocabulary. It consists of a list of terms for each of which an opposite is to be given. Sometimes, but rarely, the term is used as synonymous with *same or opposites test.*

Overlapping. This term is employed to describe the relative positions of two distributions on the same scale of measurement. Overlapping is usually measured and stated in terms of the proportion or per cent of one distribution which extends beyond the median or occasionally some other point of the other distribution with which it is being compared. For example, if the median score of a group of fifth-grade pupils on a certain test is 65, the per cent of fourth-grade pupils who score above 65 is said to be the overlapping of the fourth grade upon the fifth as regards that particular test. Overlapping is most commonly determined in connection with grade and age groups.—Odell, Educational Statistics, p. 286f.

P. One of the two common abbreviations for *percentile.*

Pantomime test. A pantomime test is the same as a non-verbal test in the narrowest sense of the term. In other words, it is a test in which no written or spoken language is used to communicate to the subjects what they are to do, but pantomine or illustrative actions by the examiner are employed for this purpose. The chief use of such tests is in measuring the abilities of persons who are unable to understand the language spoken by the examiner.

Parallel group. In the two-group or equivalent-group method of experimentation the groups concerned are sometimes spoken of as parallel groups. See *equivalent group.*

Part. The most frequent use of this term is to apply to a portion of a test or a test of a series which is intended for use in one or more grades, the other portions or tests each being intended for use in other grades or combinations thereof. Thus Part 1 of a test may be for use in Grades III and IV, Part 2 in Grades V and VI, and Part 3 in Grades VII and VIII. Occasionally the term part is used in some other sense to signify a portion of a test or a test of a series that covers different content or is in different form from the other portion thereof.

Partial correlation. Partial correlation is a method of correlation involving three or more variables in which that portion of the correlation between two of them which is not due to or common with the others included, is determined. In other words, the influence of all the variables except two is held constant or eliminated and the correlation between those two determined. Partial correlation is practically always expressed in terms of the coefficient of partial correlation, which is calculated from ordinary product-moment coefficients of correlation. See *coefficient of positive correlation, correlation.*—Odell, Educational Statistics, p. 245f. Otis, p. 230f.

P. E. Abbreviation for *probable error.* A subscript is frequently employed to indicate the situation or derived measure to which the probable error refers. Thus the subscript M. is used to denote the probable error of the mean, Md. that of the median, r that of the coefficient of correlation, and so on.

P. E.$_{est.}$. Abbreviation for *probable error of estimate.*

P. E.$_{meas.}$. Abbreviation for *probable error of measurement.*

Per. Abbreviation for *percentile.*

Percentile (Per. or P.). The percentiles are the points which divide the total number of cases contained in a frequency distribution

into 100 equal parts; that is, into 100 parts each of which contains the same number of cases. To illustrate, 5 per cent of all the cases in a given distribution lie at or below the fifth percentile and 95 per cent at or above that point, 22 per cent lie at or below the twenty-second percentile and 78 per cent at or above that point, and so on. The percentile is the smallest unit of division ordinarily employed in connection with frequency distributions.—Kelley, p. 185f. Odell, Educational Statistics, p. 111f.

Percentile curve. Synonymous with *ogive*.

Percentile norm. Although the standard method of stating norms is in terms of the median, which is the same as the fiftieth percentile, this is not infrequently supplemented by a statement of other points in the distribution. Sometimes the scores corresponding to the tenth, twentieth, and every successive tenth percentile are given and sometimes those at other percentile points. The value of such norms is that one can compare not merely the median or average achievement of a class with them, but also the achievement of pupils near the bottom, top, or other points in the distribution.—Ruch and Stoddard, p. 347f.

Percentile rank. Synonymous with *percentile score*.

Percentile score. A percentile score is a statement of a pupil's score in terms of his relative or percentile position in the distribution of scores of the whole group to which he belongs. A percentile score of a given amount, as, for example, 66, means that his score is equal to or better than the scores of the given per cent, in this case 66, of the pupils in the group. For the comparison of scores made by the same pupil on different tests or by different pupils, percentile scores are often very useful.—Monroe, Theory, p. 154f. Otis, p. 26f., 95f., 118f.

Performance. A pupil's performance is what he does. On group tests his performance is always or practically always written and the same is true for some individual tests. To be useful for testing purposes it must be such that a competent observer or scorer can easily observe it. Performance, what a pupil does, is to be distinguished from ability or capacity, what he might or is able to do.

Performance test or scale. A performance test or scale is composed of exercises which require the subject to react to problems presented in the form of concrete objects rather than of words. Instructions may be either verbal or pantomime. Thus a performance test is a variety of non-verbal test. Indeed, the two terms are sometimes used interchangeably, but in its broader sense the non-verbal test is more inclusive than the performance test.—Freeman, p. 158f.

Personal equation. It has been discovered that in measurements involving observation there tend to be constant errors present in the cases of all series of observations and that the amounts of these errors differ with different observers. This difference in the amount of error has been called the personal equation. See *subjective.*—Freeman, p. 32f.

Point scale. In a broad sense a point scale may be said to be any scale which makes use of scores computed in terms of points. The expression has, however, been generally limited to apply to general intelligence scales which are scored in terms of points as contrasted with those scored in terms of months or years of mental age. Ordinarily age norms are given in connection with such scales so that any obtained point score may be transmuted into a corresponding mental age.—Freeman, p. 131f.

Point score. A point score is the score yielded directly by a test. It may be in terms of exercises done correctly, exercises attempted, level of difficulty reached, and so forth. It is only by chance that point scores upon two or more different tests have the same meaning with regard to the amount of achievement or ability which they represent or indicate. In many cases provision is made for turning point scores into derived scores of various sorts. See *derived score.*—Freeman, p. 265.

Positive correlation. The correlation or relationship between two variables or sets of paired measures is called positive when there is a tendency for large measures in one series to be associated with large measures in the other and vice versa. See *correlation, negative correlation.*

Power test. A scaled test—that is, a test arranged in order of increasing difficulty of exercises which yields only a difficulty score—is called a power test. Such an instrument measures the power or ability of pupils to do increasingly difficult exercises of the same kind, hence the name. Sometimes the term is used as entirely synonymous with scaled test regardless of the method of scoring.—Kelley, p. 31.

Practice effect. Practice effect refers to the increase of the scores of one trial over those yielded by a preceding trial of the same test when there has been no coaching between the two administrations of the test. The term is commonly used to refer to the average increase of the scores of a group of pupils, but sometimes in connection with the increase between the scores of an individual pupil. Through becoming acquainted with testing procedure and the nature of the exercises pupils tend to make higher scores on the second trial than on

the first, still higher on the third than on the second, and so on. In general, however, the increase from the first trial to the second is much greater than that from the second to the third. This tendency continues, until after perhaps the fourth or fifth trial there is often very little or no further increase. Also the increase even from the first to the second trial is much less if pupils are used to taking tests of the same general character than if they are not. The practice effect between two trials of a test tends to be approximately the same for all pupils in the group and, therefore, constitutes a constant error. Data from a number of tests indicate that the average increase due to practice effect between the first and second trials is about 10 per cent of the first trial scores, that between the second and third trials it is usually less than 5 per cent, and that between the fourth and fifth trials it is rarely much over 1 per cent.—Monroe, Theory, p. 167f. Otis, p. 264f.

Practice test. This expression is used in two senses. In one it is synonymous with preliminary test or fore exercise. In the other it refers to a test which has as its function giving pupils practice in the abilities covered rather than measuring their achievements thereon. Such practice tests are most common in arithmetic, but also exist in algebra, language, and other subjects. Usually a rather large number of them are included in one series.

Preliminary test. Synonymous with *fore exercise.*

Principle. Principles include laws, rules, truths and certain other important statements. In other words, a principle may be thought of as a statement or criterion, usually generalized, by which the truth or validity of a proposed plan, a suggested theory, or a tentative conclusion, may be tested.

Probable error (P. E.). The term probable error should be limited in use to apply to the median deviation when used as a measure of the errors present in data of any sort. It is also frequently but improperly used as completely synonymous with median deviation. In either usage half of the deviations or errors in a normal distribution are less than the probable error and the other half are greater. In other words, the chances are even or one to one that any particular error is greater or less than the probable error. Similar statements involving, of course, different chances or proportions can be made concerning errors greater and less than 2 P. E., 3 P. E., and so on. In educational work the probable error is the most commonly used measure of errors. It is ordinarily assumed that errors form a normal distribution and, therefore, that the same interpretation of the probable error applies in all cases. Usually the approximation to a normal dis-

tribution is close enough to justify this assumption. A subscript is frequently employed with the abbreviation for the probable error to indicate the measure to which it belongs or the situation to which it applies. Thus P. $E._M$ refers to the probable error of the mean, P. $E._Q$ to that of the quartile deviation, and so forth. See *median deviation.*—Odell, Educational Statistics, p. 221f. Odell, Interpretation, p. 9f. Otis, p. 256f.

Probable error of estimate (P. $E._{est.}$). This is merely the probable error applied to errors of estimate. P. $E._{est} = .6745\,\sigma\sqrt{1 - r^2}$.—Kelley, p. 171f. Monroe, Theory, p. 348f. Odell, Educational Statistics, p. 230f.

Probable error of measurement (P. $E._{meas.}$). This refers to the use of the probable error in connection with errors of measurement. It is derived from the probable error of estimate. There are several formulae of which the most common is P. $E._{meas} = .6745\,\sigma\sqrt{1 - r}$.—Kelley, p. 171f. Monroe, Theory, p. 207f., 354. Odell, Educational Statistics, p. 230f.

Problem. In educational research the term problem is used to designate the question or questions to which answers are sought. It may be expressed by a declarative statement of the purpose of the investigation as a hypothesis to be proven or may be definitely in question form. In case the latter form is not used, the question or questions to be answered are implied.

Product-moment correlation. This name is given to the usual method of computing the coefficient of correlation, a method which owes its extended use to Karl Pearson. For a small number of cases, perhaps less than 25 or 30, the data are usually arranged in two columns, the corresponding entries in which constitute a pair of measures, whereas for larger numbers of cases a correlation or double-entry table is almost always used. The formula used in product-moment correlation compares the deviations of the corresponding pairs of measures from their means with the standard deviations of the two distributions and thus yields the coefficient of correlation. Its general form is $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$ or $r = \frac{\Sigma xy}{N\sigma_x \cdot \sigma_y}$ See *coefficient of correlation, correlation.*—Odell, Educational Statistics, p. 150f.

Prognostic test. A prognostic test is one which has for its function the prediction or prognosis of a pupil's status at some time in the future. Such a prediction is based upon the pupil's performance at the present. All, or practically all, tests have some prognostic value,

but those which have been devised especially for this purpose are in general more valid than those not so intended. The tests used for prognostic purposes may be intelligence tests, achievement tests, or tests which strictly speaking belong under neither of these classifications.—Monroe, Theory, p. 223. Ruch and Stoddard, p. 39f. Symonds, p. 363f.

Psychometric. The term psychometric refers to the measurement of mentality in its broadest sense; that is, including general intelligence, ability in specific subjects, emotional qualities, and so forth.

Q. Abbreviation for *quartile deviation.*

Q_1. Abbreviation for *first* or *lower quartile.*

Q_2. Abbreviation for *second quartile* (rarely used).

Q_3. Abbreviation for *third* or *upper quartile.*

Quality. One of the three dimensions concerned in measuring pupils' performances is quality. Sometimes this characteristic is described in terms of per cent of exercises done correctly. In such cases quality is synonymous with accuracy. Certain types of performances, such as handwriting and drawing, cannot be classified as either right or wrong. In such instances quality may be defined as merit and is described in terms of a quality scale with which the specimens produced by the pupils are compared. See *accuracy, dimensions.*—Monroe, Theory, p. 108f.

Quality scale. A quality scale is a scale composed of a set of samples or specimens arranged in order of merit. Pupils' performances are compared with the specimens or steps on such a scale and rated by determining the ones which they most resemble. Such scales are used in cases in which pupil performances cannot be rated as definitely right or wrong. Handwriting, English composition, and drawing are the three subjects in which quality scales are most widely used.—Monroe, Theory, p. 108f.

Quantitative method (or methods). Synonymous with *statistical method (or methods).*

Quartile (Q with subscript 1, 2 or 3). The quartiles are the points which divide the total number of cases in a frequency distribution into four equal parts; that is, into four parts each of which contains the same number of cases. Thus one-fourth of all the cases lie at or below the first quartile and three-fourths at or above it, two-fourths at or below the second quartile and two-fourths at or above it, and three-fourths at or below the third quartile and one-fourth at or above it. The first and third quartiles are very commonly given along

with the median, which is the name applied to the second quartile, in describing a distribution. The term quartile is also sometimes applied to one of the four divisions formed by the points just mentioned. See *first quartile, second quartile, third quartile.*—Odell, Educational Statistics, p. 111f.

Quartile deviation (Q.). One of the most common measures of deviation or dispersion is the quartile deviation, also sometimes called the semi-interquartile range. It is found by taking half of the distance from the first to the third quartile or, in other words, by taking half of the distance which includes the middle 50 per cent of the cases. In formula form, $Q = \frac{Q_3 - Q_1}{2}$. In a normal distribution it becomes the same as the median deviation, but it is only by chance that this is exactly true in a distribution which is not normal.—Odell, Educational Statistics, p. 120f. Rugg, p. 155f.

Questionnaire. The questionnaire or question blank has come to be a very much used and very much abused device for gathering educational data. It consists of a more or less formal list of questions, copies of which are sent to a number of persons with the request that they fill in the answers and return. Questionnaires run all the way from only two or three questions to several hundred and are sent to from a very few persons up to hundreds and occasionally even thousands. They also vary with reference to the types of questions asked. Some call for facts in the possession of the recipient or easily obtainable by him. Others require him to collect information and perhaps even to make calculations. Still a third type consists of questions asking for expressions of opinion. Questionnaires are least objectionable when they are of the first sort; that is, when they call for simple facts in the possession of the recipient. The questionnaire method, however, has been very much abused by being frequently employed when the data desired are already available in published form or are otherwise accessible to the investigator. Unless the need is urgent, a questionnaire should not require the recipients to collect data, and it should never ask them to make calculations. When expressions of opinion are sought, those to whom it is sent should be competent.—Rugg, p. 40f.

Quotient score. A quotient score is one which expresses a pupil's performance in comparison with his supposed ability to perform, ordinarily measured by either his general intelligence or his age. See *achievement quotient, educational quotient, intelligence quotient, subject quotient.*—Freeman, p. 285f.

R. This symbol is the abbreviation for two different expressions or measures used in connection with correlation. One is the coefficient of multiple correlation. When thus used R is followed by subscripts all but the first of which are either enclosed in parentheses or follow a dot, thus: $R_{1(23\ldots n)}$, or $R_{1\cdot 23\ldots n}$. The first subscript in this notation denotes the one variable which is correlated with the others in combination and of course the subscripts within the parenthesis or after the dot indicate those variables which form the combination. In its other usage R is the abbreviation for one of the coefficients of *rank correlation* rather commonly used. In this sense it rarely has a subscript.

r. This is the very commonly used abbreviation for the ordinary or product-moment *coefficient of correlation.* It is also used for the *coefficient of partial correlation,* in which case it is practically always followed by two subscripts, which indicate the two variables correlated, then a dot and other subscripts, which indicate the variables eliminated or held constant, thus: $r_{12\cdot 34\ldots n}$.

Random error. Synonymous with *variable error.*

Random sample. A sample is said to be random when it has been selected from the total population or group which it is to represent without any bias entering into its selection. In other words, a random sample is one selected in a purely chance manner. The accuracy or reliability with which a random sample represents the entire group—that is, how nearly it is typical of the whole group—is shown by any one of several measures of errors of sampling. See *error of sampling, sampling.*

Range. The range of a series of scores or other measures is the distance from the lowest to the highest measure. Thus the range of a group of percentile marks of which the lowest is 62 per cent and the highest 99 per cent, is 37.—Odell, Educational Statistics, p. 119f., 140. Rugg, p. 154f.

Rank correlation. In cases wherein comparatively small groups of individuals, usually not over 25 or 30, are concerned, it is very common to determine relationship by computing rank correlation rather than product-moment correlation. In so doing the ranks of the various individuals concerned are dealt with rather than their exact scores. The chief reason why rank correlation is used is that for such small numbers its computation is decidedly easier than that involved in product-moment correlation. When the number of cases becomes large, however, this is no longer true. There are two common methods of computing rank correlation, neither of which is quite as reliable as

product-moment correlation, although the difference is not great. The formula used in one method is $\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$ and that in the other, $R = 1 - \frac{6\Sigma g}{N^2-1}$. The coefficients of rank correlation obtained from these formulae may be, and usually are, turned into approximate equivalents of coefficients of product-moment correlation. See *correlation.*—Kelley, p. 189f. Odell, Educational Statistics, p. 201f. Otis, p. 206f.

Rate score. A rate score is a measure of a pupil's rate of work. It is usually expressed in terms of the number of exercises or other units of work done within a certain time. Sometimes all those attempted are counted, sometimes only those correctly answered. A rate score may also be expressed in terms of the amount of time used by a pupil to complete a specified amount of work, but this is not so common as the preceding method.

Rate test. A rate test is one which yields a rate score. It may yield other scores also, but must yield a rate score unaffected by the other dimensions of pupil performance.—Monroe, Theory, p. 63f., 107f.

Ratio score. A ratio score is similar to a quotient score although the two cannot be said to be absolutely synonymous. The term ratio score is rarely used, but when employed is usually applied to the quotient obtained by dividing an achievement score expressed in terms of age by mental age. See *quotient score.*

Ratio of correlation (eta, η). The ratio of correlation is the only commonly used index of curvilinear correlation or relationship. It must always be equal to or greater than the coefficient of correlation, being equal to it in case the relationship is rectilinear and being increasingly greater than it the more curvilinear the relationship is. It is always positive, ranging from +1.00 down to zero, and thus does not indicate whether the relationship is positive or negative. There are two ratios of correlation for each correlation table. One of these measures the curvilinear correlation of the variable shown on the horizontal scale on the one shown on the vertical scale. The other measures that of the variable shown on the vertical scale on the one represented on the horizontal scale. Using X and Y for the two variables, the formula for the ration of X on Y is $\eta_{xy} = \frac{\sqrt{\Sigma\frac{(\Sigma x)^2}{f}}{N} - c_x^2}{\sigma_x}$, and that for Y on

X is $\eta_{yx} = \frac{\sqrt{\frac{\Sigma\frac{(\Sigma y)^2}{f}}{N} - c_y^2}}{\sigma_y}$.—Odell, Educational Statistics, p. 207f.

Raw score. A raw score is the numerical expression or description of an individual's performance in terms of the unit used in the construction of the scale or in scoring the test. In order to have significance a raw score must be transmuted into a comparative or relative measure, or be compared with a norm or standard, which amounts to practically the same thing.—Freeman, p. 263f.

Recall test. Synonymous with *single-answer test.*

Recognition test. Synonymous with *multiple-answer test.*

Rectilinear relationship. The relationship between two variables is said to be rectilinear or straight-line when a graphic representation thereof is a straight line or approaches it more nearly than any other common geometrical curve. The rectilinear relationship between two or more variables is usually summarized by the coefficient of correlation, an expression which measures this type of relationship only. For purposes of predicting or estimating scores, and so forth, the regression coefficients and equations are the measures of rectilinear relationship commonly employed.

Regression. See *coefficient of regression, regression equation.*

Regression equation. For each correlation table showing the relationship of two variables there are two regression equations. One of these expresses the most probable or likely value of the first variable in terms of the second and the other that of the second in terms of the first. Thus these equations furnish the best means of predicting values of one variable when those of the other are known. The most convenient form of the formula for the regression of one variable, X, upon the other, Y, is probably as follows: $X = r\frac{\sigma_x}{\sigma_y}Y + M_x - r\frac{\sigma_x}{\sigma_y}M_y$. In connection with the correlation of three or more variables, partial or multiple regression equations may also be found by means of which the most probable value of one variable may be predicted in terms of all the others concerned. The regression equations are rectilinear; that is, they assume straightline relationship. See *coefficient of regression.*—Odell, Educational Statistics, p. 189f. Rugg, p. 248f., 254f.

Reliability. See *reliable.*

Reliable. A test or measuring instrument is reliable to the degree to which a second application of the test yields scores equivalent to

those obtained from the first application. This includes both the use of the identical test on two occasions and also of equivalent forms of the same test. In either case it will be found that some pupils make higher scores and others lower upon the second trial than on the first. Most of these differences are due to the presence of variable or accidental errors in both sets of scores. The reliability of a test is expressed in terms of a numerical coefficient or index which indicates the size of these variable errors. Constant errors do not affect reliability.—Kelley, p. 33, 35f. Monroe, Theory, p. 201f. Ruch and Stoddard, p. 51f., 355f.

Research. Research may be defined as a method of studying problems whose solutions are to be derived partly or wholly from facts. The facts dealt with in research may be statements of opinion, historical facts, those contained in records and reports, the results of tests, answers to questionnaires, experimental data of any sort, and so forth. The final purpose of educational research is to ascertain principles and develop procedures for use in the field of education; therefore it should conclude by formulating principles or procedures. The mere collection and tabulation of facts is not research though it may be preliminary to it or even a part thereof.—Monroe and Engelhart, p. 7f.

Rho (ρ). Abbreviation for one of the common coefficients of *rank correlation.*

Right-minus-wrong formula. This refers to the formula commonly and preferably used in scoring alternative tests. According to it a pupil's score consists of the number of right answers minus the number of wrong answers. It is also sometimes used in connection with multiple-answer tests involving more than two possibilities. The generalized form of the formula which applies to all multiple-answer tests is: $\text{Score} = R - \frac{W}{N-1}$. In this formula R equals the number of right answers, W the number of wrong answers, and N the number of suggested answers in each exercise.—Odell, Objective Measurement, p. 16.

Root-mean-square deviation. This term is applied to measures of deviation or variability based upon the squares of the deviations. The only one of these measures commonly used is the standard deviation. Frequently the term is used as exactly synonymous with standard deviation but it should be followed by the qualifying phrase "from the mean" if this is done. See *standard deviation.*

Rotation method. This is a method of arranging or organizing groups of pupils for experimentation. It involves the use of two or

more groups in which the experimental factors are rotated so as to yield a more nearly equivalent basis of comparison.—McCall, How to Experiment, p. 19f., 31f.

S. A. Abbreviation for *subject age.*

Same or opposites test. This is a variety of objective test sometimes used as a form of the new examination and also in standardized tests in which a number of pairs of words or other expressions are given and the pupils are to indicate whether those in each pair mean the same or the opposite.—Odell, Objective Measurement, p. 19f.

Sampling. In educational research it is very commonly desired to study a group so large that all members of the group cannot be included. It therefore becomes necessary to resort to sampling; that is, to the selection of a portion or sample of the whole group with which it is desired to deal. This sample is then studied and the results obtained considered as applying to the whole group. The sample selected should be so chosen that no bias enters into its selection and should be large enough to yield fairly reliable results. How reliable these results are can ordinarily be determined by measuring errors of sampling. See *error of sampling, random sample.*

Scale. The word scale is used in two somewhat different yet related senses. In the most restricted of these it designates that portion of a measuring instrument which is used in describing a pupil's performance as contrasted with that portion which secures the pupil's performance. In the case of some of our measuring instruments, such as composition and handwriting scales, the scale itself is the conspicous feature and the procedure which must be followed in order to secure pupil performances is not a part of the scale. In the case of other measuring instruments, such as common standardized tests in arithmetic and spelling, the scale is less obvious, the test portion of the instrument being prominent. There must be in the case of every measuring instrument, however, some scale composed of units in terms of which pupils' performances are described just as a scale for measuring height must be in terms of meters, feet, inches, or some other unit, one for weight in terms of pounds, ounces, or something else, and so on. In its second sense the word scale is used as synonymous with scaled test. It should perhaps also be mentioned that sometimes scale is incorrectly and carelessly used as synonymous with *test.*—Monroe, Theory, p. 15f., 20f., 106.

Scaled test. A scaled test is one in which the exercises are arranged in order of increasing difficulty. It is a frequent and desirable,

but not necessary, feature that the increase in difficulty from one exercise to the next be approximately constant throughout the scale. See *power test.*—Monroe, Theory, p. 62, 73f., 78f., 89f., 118f.

Scatter diagram. Synonymous with *correlation graph.*

School survey. This term is used to describe a study or investigation of a city, state, or other school system, or in some cases of a single school, which attempts to evaluate the general efficiency thereof and to point out needed changes and improvements. Such a survey ordinarily deals with the building program, finances, qualifications and salaries of teachers, pupil achievement, general administration and organization, methods of supervision and teaching, the curriculum, and various other factors. Sometimes a survey is limited in scope, dealing with only one or a few of the matters mentioned. Thus there may be a building survey, a financial survey, a survey of teaching personnel, and so forth.

Scientific. Strictly speaking, anything based upon facts is scientific. For the field of educational research an investigator may be called scientific when he knows his data and uses them with a complete recognition of any imperfections that may exist either in them or in his procedures. The significance of this statement becomes more fully apparent when we realize that in educational research the data dealt with are seldom, if ever, perfect.—Monroe and Engelhart, p. 49f.

Score. A pupil's score is a description of his performance. As distinguished from a mark it is a description in terms of the scale of units used in connection with the given measuring instrument and not in terms of the marking system employed in the school.—Monroe, DeVoss, and Kelly, p. 417f.

S. D. One of the two abbreviations for *standard deviation.* See *sigma* (σ).

Second quartile (Q_2). Synonymous with *median*, therefore the expression is rarely used.

Selection of exercises. In the construction of educational tests it is usual to secure a large number of exercises and select from this number those to be used in the final test. Such a selection may be in accord with any one or any combination of three criteria or methods, or it may be without the use of any definite criteria. These three are statistical selection, agreement with educational objectives, and suitableness for testing purposes as determined by trial. If no definite criterion is used the selection is said to be arbitrary.—Monroe, Theory, p. 89f. Ruch and Stoddard, p. 304f. Symonds, p. 279f.

Selection test. This term is sometimes applied to any one of several varieties of objective tests. Among these are the matching test, the test which calls for a rearrangement of items in the correct order, certain varieties of multiple-answer tests, and so forth.—Russell, p. 89f.

Self-correlation. This refers to correlation employed for the purpose of measuring reliability. See *correlation, reliable.*

Semi-interquartile range (Q). Synonymous with *quartile deviation.*

Short-answer test. Synonymous with *new examination.*

Sigma (Σ). The capital sigma is used as the symbol of summation; that is, it indicates that various values of the variable referred to are to be summed or added. For example, the expression ΣX means that all values of the variable X are to be summed.

Sigma (σ). The most common abbreviation for the *standard deviation* or *standard error.* A subscript is frequently employed with the abbreviation for the standard deviation to indicate the measure to which it belongs or the situation to which it applies. Thus σ_M denotes the standard deviation or error of the mean, σ_b that of the coefficient of regression, $\sigma_{est.}$ the standard error of estimate, and so forth.

$\sigma_{est.}$. Abbreviation for *standard error of estimate.*

$\sigma_{meas.}$. Abbreviation for *standard error of measurement.*

Significance. In a technical statistical sense a measure or difference is said to be significant when by comparison with its standard or probable error or some other measure of reliability it is apparent that it is fairly reliable. The most common meaning of significance has to do with sampling; that is, with whether or not the errors resulting from using only a sample are so great as to destroy the significance of the derived measures or conclusions. The question of significance also rather often arises in connection with the effect of errors, particularly variable errors, upon derived measures. If a measure or difference is two times its standard error or three times its probable error, it is ordinarily considered significant, though sometimes this ratio is raised to three times the standard error and four or five times the probable error.—Odell, Educational Statistics, p. 221f.

Similarities test. This is a variety of the multiple-answer or association test in which the one or more of several given terms most like one or more other given terms is to be indicated.

Single-answer test. This is a variety of the new examination which consists of questions so phrased that the answer to each is a

single word. It is ordinarily understood also that the questions are such that there is only one possible correct answer.—Odell, Objective Measurement, p. 9. Ruch and Stoddard, p. 267, 272.

Sk. Abbreviation for *skewness.*

Skew (or skewed) distribution. A skew distribution or frequency curve may be thought of as a normal distribution or curve which has been pushed or pulled out in one direction so that one extreme is further from the central tendency than the other. If it has been stretched out so that the end of the distribution at which the largest measures are located is further from the central tendency, the skewness is said to be positive or plus. If the lower end is further from the central tendency, it is said to be negative or minus. The most common formulae for measuring skewness are $\text{sk.} = \frac{3(\text{M.} - \text{Md.})}{\sigma}$ and $\text{sk.} = \frac{Q_3 + Q_1 - 2\text{Md.}}{Q}$.—Odell, Educational Statistics, p. 59f., 281f. Rugg, p. 178f. Russell, p. 215f.

Skewness. See *skew distribution.*

Smoothed curve. In cases in which the data are too few to be truly representative and therefore show irregularities not typical of the whole group being studied, they are smoothed—that is, rounded off—to approximate the distribution that would supposedly be obtained if the sample were adequate in size. The most common method of smoothing consists in substituting for each frequency a new frequency which is the average of the original one and a given number of adjacent frequencies half of which lie on each side of it. The usual number of such adjacent frequencies taken is two, one on each side of the original frequency.—Odell, Educational Statistics, p. 45f. Rugg, p. 182f.

Social age. Just as general intelligence is frequently stated in terms of mental age and achievement in terms of achievement or subject age, so social development or maturity is sometimes stated in terms of social age. A social age of a given amount such, for example, as twelve years and six months, means that the individual so rated has the maturity that is typical or average for children twelve years and six months old.

Speed test. Synonymous with *rate test.*

Spiral test. A spiral test is a cycle test so arranged that there is an increase in difficulty in successive sub-tests or exercises. Thus in arithmetic such a test may first have easy exercises in addition followed

by easy ones in subtraction, multiplication and division, then more difficult ones in each of these fundamentals, then still more difficult ones, and so on. Most spiral tests are not entirely regular or uniform in increase in difficulty and in rotation of types of exercises. See *cycle test.* —Monroe, Theory, p. 63, 74f.

S. Q. Abbreviation for *subject quotient.*

S. R. Abbreviation for *subject ratio.*

Standard. A standard is a statement of the goal or objective which pupils should reach in their performance at a certain time. It is usually stated as an age or grade standard. Standards may be based upon norms but differ from them in that they represent goals of attainment rather than average actual attainment.—Symonds, p. 260f.

Standard deviation (σ. or S. D.). The standard deviation is one of the two or three most common measures of deviation or variability used. It is based upon the squares of the actual deviations and is always found about the mean. In a normal distribution or curve it represents the distance from the mean to the point of inflection; that is, the point at which the slope of the curve changes from an angle of more than 45° with the base line to one of less than that amount. Furthermore in a normal distribution a distance of one standard deviation on each side of the mean includes 34.13 per cent of the area of the curve or, in other words, of the number of cases. Therefore 68.27 per cent of the cases in a normal distribution lie not more than one standard deviation from the mean. The simple formula for the standard deviation is $\sigma = \sqrt{\frac{\Sigma x}{N}}$.—Kelley, p. 154f. Odell, Educational Statistics, p. 128f. Rugg, p. 167f.

Standard error (σ). This is merely the standard deviation when used as a measure of errors.

Standard error of estimate ($\sigma_{est.}$). This refers to the standard error when used as a measure of errors of estimate. $\sigma_{est.} = \sigma\sqrt{1 - r^2}$. —Monroe, Theory, p. 348f. Odell, Educational Statistics, p. 230f.

Standard error of measurement ($\sigma_{meas.}$). This is merely the standard error used to measure errors of measurement. It is derived from the standard error of estimate. $\sigma_{meas.} = \sigma\sqrt{1 - r}$.—Monroe, Theory, p. 207f. Odell, Educational Statistics, p. 230f.

Standard test. This expression is sometimes used as synonymous with *standardized test* in the broader sense of the latter term.

Standard unit. A standard unit is one which is understood in the same way; that is, whose magnitude is known, by all persons com-

petent to deal with it. Examples of such units are: a foot, a bushel, a year. A unit may be made standard by use, by authority, or otherwise.—Monroe, Theory, p. 17.

Standardized test. In the strictest sense of the term a test is standardized when norms based upon a sufficient number of individuals have been determined for it. In this sense there are no requirements to be fulfilled as to the form and structure of the test, the selection of exercises contained therein, the administration, or the scoring. In common usage, however, the expression standardized test is understood to have a somewhat broader meaning and to refer to a test which not only has satisfactory norms, but also has been devised so that it yields relatively objective scores, has such directions for administration as to secure practical uniformity, and on the whole meets the criteria of a satisfactory test fairly well.—Monroe, DeVoss, and Kelly, p. 12.

Statistical method (or methods). In a broad sense this refers to any method of research or investigation which involves even the simplest mathematical operations. The expression is, however, usually employed in a more limited sense to refer to procedure which involves somewhat elaborate tabulation of data and statistical treatment of the results.—Monroe and Engelhart, p. 42f.

Statistical selection of exercises. One of the methods of selecting the exercises to be included in a test from the large number usually collected is known as the method of statistical selection. According to this the per cent of correct responses for each exercise is determined and from these data the difficulty of each computed. The exercises then selected are those whose degrees of difficulty are appropriate to the structure of the desired test. It is usually desired either to secure exercises all of which are of approximately the same difficulty, or which are of increasing difficulty beginning with relatively easy and running to relatively difficult and with approximately constant intervals between each pair of adjacent exercises.—Monroe, Theory, p. 89f.

Subject age (S. A.). Synonymous with achievement age, except that subject age is used only in connection with single subjects, never with an average age in several subjects. See *achievement age, educational age.*

Subjective. A measuring instrument is said to be subjective when different results are secured by different persons, or by the same person at different times, using it to measure the same thing. The cause of subjectivity may be in the giving of the test to the pupils or in the scoring of their responses. In the latter case the scoring is said to be subjective, which means that different persons or the same person

at different times tend to assign different scores to the same responses. Thus subjective is the opposite of objective. Practically no test is either entirely subjective or entirely lacking in subjectivity, so that the term is commonly used in a relative sense and a test which possesses a high degree of subjectivity is said to be subjective.—Monroe, Theory, p. 26f.

Subjectivity. See *subjective.*

Subject-matter test. Synonymous with *achievement test.*

Subject quotient (S. Q.). A subject quotient is found in the same general manner as an achievement quotient; that is, by dividing a pupil's score expressed in terms of subject age by his chronological age. Thus S. Q. $= \frac{\text{S. A.}}{\text{C. A.}}$. The expression is used only in connection with separate subjects and not with combined or composite scores. See *achievement quotient, educational quotient.*

Subject ratio (S. R.). This expression, which is very rarely used, refers to the quotient obtained by dividing a pupil's score in a particular subject expressed in terms of subject age by his mental age. It is, therefore, synonymous with the achievement quotient in the ordinary sense of the latter, except that it is never used in connection with a composite or combined score. See *achievement quotient.*

Sub-test. A sub-test is one of the major divisions of a test or measuring instrument. All the exercises within each sub-test are of the same general form or type. Many tests are not divided into sub-tests and hence may be thought of as consisting of just one sub-test.

Survey. Synonymous with *school survey.*

Survey test. Synonymous with *general survey test.*

Table of double entry. Synonymous with *correlation table.*

10-90 percentile range (D). The distance between the tenth and the ninetieth percentiles has been suggested and used as a measure of deviation or variability. In formula form, $D = P_{90} - P_{10}$.—Odell, Educational Statistics, p. 122f.

Test. The word test is used in a general sense to designate any type of instrument for measuring mental capacity or ability of any sort. In this usage it includes instruments which have been designated tests by their authors and likewise those which have been called scales, as well as ordinary examinations. In a restricted sense it refers to the portion of a measuring instrument that is employed to secure pupil performances, as distinguished from a scale, which is the portion used to measure the performances when secured. In the case of some of

our measuring instruments the test feature is much more prominent, whereas in the case of others the scale feature is so. Still a third usage is sometimes found. According to this the word test is used to include all measuring instruments which present exercises or questions to which the pupils respond directly and to which the responses may in general be scored as right or wrong in contrast to those which consist of sets of specimens or samples with which pupils' performances are compared. This usage is, of course, a slight modification of the second meaning given.

Third quartile (Q_3.). The third quartile is that point on the scale of measurement used in connection with any distribution or series of measures at or below which three-fourths and at or above which one-fourth of the measures fall. Its formula is $Q_3 = 1 + \frac{\frac{3N}{4} - S}{f.}$. See *quartile.*—Odell, Educational Statistics, p. 111f.

Timed test. A timed test is for practical purposes synonymous with a rate test. Sometimes tests, usually scaled or power tests, have time limits given which are long enough that practically all pupils are able to advance as far along the scale as their ability permits before time is called. In such cases they should not be described as timed. In the case of some timed tests in which the limit is really effective, however, the method of describing pupil performances is such that no separate and distinct rate score is yielded.

Traditional examination. This term has come to be frequently applied to examinations of the type commonly used until at least very recently and probably yet much more common than any other variety. Such examinations consist of exercises which require pupils to discuss, summarize, outline, criticise, compare, reorganize, evaluate, state, show, analyze, and so forth. The term is used in contrast to new examination and is, therefore, generally understood to include tests or examinations which are relatively subjective and require a considerable amount of writing on the part of pupils.—Ruch and Stoddard, p. 252f. Russell, p. 166f.

Transmuted score. A transmuted score is one which has been changed from its original form or numerical value as a point score yielded directly by a test into an equivalent score on some other basis. See *derived score, transmutation of scores.*

Transmution of scores. The transmutation or changing of scores generally refers to the changing of point scores—that is, scores yielded directly by a test or scale—into ratings of some other sort, such as age

scores, T-scores, school marks, and so forth. Sometimes also point scores on one or more tests are transmuted so as to be equivalent to scores on another test or perhaps all are changed to some common basis for purposes of comparing, combining, averaging, or other computation.—Monroe, Theory, p. 211f. Odell, Educational Statistics, p. 196f., 295f. Otis, p. 119f.

True-false test. An alternative test which consists of a number of statements the truth or falsity of which is to be indicated by those being tested, is called a true-false test. This form of exercise is rather commonly used in connection with new-type examinations and standardized tests.—Odell, Objective Measurement, p. 10f. Ruch and Stoddard, p. 268, 275. Russell, p. 28f.

True score. A pupil's true score may be defined as the average of an infinite number of measurements of the characteristic being measured. These measurements should be made under the same conditions. It is, of course, impossible to fulfill either the ideal of an infinite number of measurements or that of the same conditions. Even though other conditions are controlled as well as possible, practice effect enters in and in general causes higher scores to be made on the second trial of the test than on the first, on the third than on the second, and so on. Therefore, in some cases an approximation to a true score is obtained which consists of the average of a fairly large number of measurements corrected as well as possible for practice effect and other differences in the testing conditions. The concept of a true score is frequently helpful even though such a score cannot actually be found and certain statistical calculations concerning true scores can be made even though the scores themselves cannot be determined.—Monroe, Theory, p. 201f.

T-scale. The T-scale, so named in honor of Terman and Thorndike, is a scale based upon the distribution of ability of an average or complete group of twelve-year-old pupils. It consists of 100 units of .1 standard deviation each and extends from five standard deviations below the mean of twelve-year-old pupil ability to five standard deviations above the mean. For pupils whose abilities are not too different from those of twelve-year-old pupils it provides a basis for derived scores which may be compared with one another though derived from different tests. A rather large number of standardized tests provide tables by which point scores may be transmuted into T-scores.—McCall, How to Measure, p. 272f. Monroe, Theory, p. 150f. Ruch and Stoddard, p. 350f.

T-score. A score given according to the T-scale.

Two-groups method. This is synonymous with the *equivalent groups method* when only two groups of pupils are employed.

Undistributed scores. In the cases of some of our measuring instruments the easiest exercises are so difficult that pupils who make scores of zero may represent a considerable range in ability. In the case of others the most difficult exercises are so easy or the time so long, or both, that a number of pupils frequently make perfect scores and thus no complete information is secured as to the extent of their abilities. Furthermore, in some tests the scale units employed are so large or the difference in difficulty between successive exercises so great that there may be considerable differences in the abilities of pupils who earn the same score. In such cases as all these it is said that the scores of the pupils whose abilities differ but who receive the same scores in so far as a given test is concerned are undistributed. See *discrimination.*

Uniform test. Synonymous with *rate test.*

Unreliability. See *reliability.*

Unreliable. See *reliable.*

Upper quartile (Q_3.). Synonymous with *third quartile.*

Valid. A measuring instrument is commonly said to be valid if it fulfills the function which it is intended or stated to perform. It may lack validity either because it is unreliable, due to subjective administration and scoring, or because it measures some other ability or abilities than its function specifies. Thus a test cannot be valid unless it is objective and reliable, but can be perfectly objective and reliable without being valid. Since few, if any, tests possess perfect validity, the term is used in a relative sense and the tests are said to be valid when they approximate validity. It has also been suggested that the term valid should be used in a more restricted sense than that just explained. In this sense it would exclude the factor of reliability. That is to say, a measuring instrument would be called valid if it performed its stated function better than any other which might be stated for it regardless of how well it did so. Thus a test might be so unreliable that little confidence could be placed in the scores obtained from it, but if they were better measures of its stated function than of anything else it would be valid.—Kelley, p. 30f. Ruch and Stoddard, p. 48f., 301f. Monroe, Theory, p. 188f.

Validation. See *valid.*

Validity. See *valid*.

Variable. As a noun the term variable is used to refer to a characteristic or trait which may exist in different amounts. To illustrate, pupils' heights differ, one pupil possessing a certain amount or degree of height, another a different degree, and so on; therefore height is a variable. Again, the quality of pupils' handwriting differs, since that of one pupil may possess a certain degree of merit, that of another pupil a different degree, and so forth; therefore quality of handwriting also is a variable. Because almost all of the traits dealt with in educational work are variable the term is very commonly used to refer to the two or more traits or characteristics which are compared, correlated, or dealt with in some other way. Variable is also used as an adjective in at least two different senses. Sometimes it is used in the same meaning as when a noun; thus any variable (noun) may be said to be variable (adjective). On other occasions it is used, most often in the phrase "variable error," as synonymous with chance or accidental.—Odell, Educational Statistics, p. 124.

Variable error. Variable errors differ for the different members of a group as contrasted with constant errors which tend to be the same for a whole group. Approximately half of the variable errors in a given group are positive and the other half negative, usually, however, a few being zero. The distinguishing characteristics of variable errors are that they differ from pupil to pupil and that ordinarily the magnitude of the variable error in the case of any given individual cannot be determined. It is, however, practically always possible to make statements as to the general size and distribution of the variable errors in a group and as to the chances that the variable error does or does not exceed a certain magnitude in the case of any particular individual. If one pupil breaks a pencil point and thereby loses a little time, if another cheats by copying from a neighbor, if a third just happens to have reviewed the material covered by a test very recently, if a fourth happens to be under par mentally and physically, the resulting differences in scores from what they would be if these peculiar conditions did not exist constitute variable errors. From the standpoint of effect upon derived measures variable errors differ from constant errors in that they do not affect measures of central tendency—that is, averages—but do tend to lower coefficients of correlation, whereas just the reverse is true of constant errors. See *constant error*. Monroe, Constant and Variable Errors.—Monroe, Theory, p. 198f., 248, 329f., 344.

Variability. Synonymous with *deviation*.

Verbal test. Sometimes all tests in which either the examiner or the subjects make use of spoken or written language are called verbal. On other occasions the term is applied only to those tests in which the subjects must respond by written or spoken language and not to those in which oral directions are given by the examiner with no verbal responses by the subjects.—Freeman, p. 257f.

Vocational guidance. This refers to the guidance or advising of individuals with regard to choosing their vocations or occupations. No hard and fast line can be drawn between it and educational guidance as much of one is frequently necessary in connection with the other.

Weighting. The determination of the proportional part to be played by each of a number of items or factors in determining a total or average score or measure is called weighting. The most frequent occasion for determining weights is in connection with the various exercises or other parts of a test or examination. If a correct response to one exercise is given a credit of three points, that to another of two, and to a third of one, the weights of these exercises are said to be respectively three, two, and one. A test in which all exercises count the same number of points, frequently one for each, is sometimes said to be unweighted, but improperly so, since the exercises are in reality equally weighted. In the cases of many standardized tests weights have been assigned in accordance with rather careful determinations of difficulty. In other standardized tests the determining factor has been the relative or supposed relative importance of the exercises. Other plans of weighting, some of which are merely modifications of the two described, have also been used. Experimental studies have shown that unless the number of items is small or the differences in weights very great, the relative scores of pupils will differ little, if all exercises or items are weighted equally, from what they will be if weights are carefully determined. In a similar fashion to that just described, weighting is also necessary in determining pupils' standings for the semester or year from their marks upon oral recitation, short quizzes, outside written work, notebooks, laboratory work, final examinations, and any other elements considered. Weighting also frequently enters into the determination of a criterion measure, in which case a number of different measures are frequently combined into one.—Freeman, p. 272f. Monroe, Theory, p. 116f. Ruch and Stoddard, p. 332f.

X, x. In dealing with situations in which two variables are concerned, such as a correlation table, the coefficient and ratio of correlation, the regression equations, and so forth, it is very common to

refer to one of them by the term X. If they are in a correlation table the one so referred to is that which has its scale upon the horizontal axis. Whenever X is used to refer to the variable itself, x is used to refer to the difference or deviation of the variable from its mean. See *correlation table, variable.*—Odell, Educational Statistics, p. 36f., 156f.

Y, y. In dealing with situations in which two variables are concerned, such as a correlation table, the coefficient and ratio of correlation, the regression equations, and so forth, it is very common to refer to one of them by the term Y. If they are in a correlation table the one so referred to is that which has its scale upon the vertical axis. Whenever Y is used to refer to the variable itself, y is used to refer to the difference or deviation of the variable from its mean. See *correlation table, variable.*—Odell, Educational Statistics, p. 36f., 156f.

Yes-no test. This is a variety of the alternative test commonly used in connection with the new examination and upon standardized tests. It consists of a series of questions to each one of which pupils are expected to respond by yes or no.—Odell, Objective Measurement, p. 9f.

Z. Abbreviation for mode.

Zero point. The zero point on any given scale is the point which means just not any of the trait or characteristic measured by that scale. In the case of most educational measuring instruments a score of zero does not represent zero ability, or, in other words, a pupil who earns a score of zero cannot be known to be located at the true zero point. This result follows from the fact that the easiest exercises on most tests are difficult enough that a pupil may have some knowledge or ability along the line tested and still not be able to respond correctly to the easiest exercise on the test. If scores on different tests are expressed in terms of a common unit they can, for some purposes at least, be added to and subtracted from one another without the determination of true zero points, but they cannot be multiplied and divided into one another unless such points have been found.—Monroe, Theory, p. 101f., 146f., 150.

BULLETIN NO. 41

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

RECONSTRUCTION OF THE SECONDARY-SCHOOL CURRICULUM: ITS MEANING AND TRENDS

By

WALTER S. MONROE
Director, Bureau of Educational Research

and

M. E. HERRIOTT
Associate, Bureau of Educational Research

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1928